

External assessment pays national dividends

A decision to expose Germany's science to international scrutiny has yielded a constructive appraisal. Although such evaluation can have its pitfalls, other countries should follow suit.

The ability to see ourselves as others see us is a gift. So bringing in foreign experts to assess the operation of two of Germany's basic research organizations — the Max Planck Society (MPS) and the Deutsche Forschungsgemeinschaft (DFG) — was in practice a generous idea of the Bund-Länder Kommission, despite initial grumblings.

The MPS and DFG are generally viewed as solid, well-funded organizations of high scientific integrity and an enviable level of political autonomy. But are they, perhaps, also slow to take up new ideas? The evaluation report delivered last week (see pages 395–396) puts some flesh on such concerns. German science tends to adopt a low-risk approach, it concludes. Although German scientists and administrators are aware of the problems it highlights, this international recognition of the issues should weaken some of the resistance to change that German reformers have encountered.

Commendably, the review exercise has raised the urgency of dealing with the poor career prospects for young scientists and employment laws that prevent German research institutions from attracting top scientists with the required salaries. Resolution of these issues lies partly in the hands of politicians. A relaxation of laws governing employment in the public sector is essential for progress on both fronts.

One obvious weakness of this particular evaluation exercise is that it was limited to a small — albeit critical — part of the German research system. A second is that it was, by definition, carried out by experts from other cultures, inevitably bringing with them their own cultural biases. Perhaps it is as a consequence of this that the evaluation report suggests that the DFG should foray into the realms of research policy by supplementing its bottom-up approach to research funding with strategic grants programmes.

This suggestion is out of tune with both the political independence of the DFG, which is a primary strength of that organization,

and the operation of the German research funding system as a whole. The federal research ministry, as well as the Helmholtz Society (the umbrella group for Germany's 16 large national research centres), run major strategic programmes, which could be expanded if necessary. The DFG should not be expected to operate like research organizations in other countries which have political as well as non-political responsibilities.

Similarly, the suggestion, however tentative, that the MPS should move away from the Harnack principle — whereby research is concentrated in institutes headed by directors endowed with research freedom and strong resources — is misplaced. That philosophy, a cornerstone of the MPS operation, places much responsibility on individuals. But the recent introduction of institute evaluations reduces the risk that those directors who do not fulfil early promise will continue to consume resources unchecked. Many countries are taking trouble to ensure that the most gifted scientists prosper through freedom and independence: the MPS provides an inspiration.

Furthermore, the suggestion that MPS resources be spread into the universities by setting up research units away from institutes is not the answer to two undeniable problems: the patchy quality of university research and the inadequate cooperation between universities and Max Planck institutes. The cure for the universities' malaise must come from within, primarily by introducing a much higher level of competition between them.

A national research system needs occasionally to be evaluated in its entirety — a task that the Organisation for Economic Co-operation and Development used to pursue but has now regrettably ceased. Such a large-scale evaluation might be timely in Germany, where more integration of its many strands of research is needed. In the meantime, that country's decision to subject itself to independent outside scrutiny should be applauded. Other scientifically developed nations should do the same. □

Policy on papers' contributors

Nature is encouraging authors of papers to say who did what.

This week sees a small step forward for transparency in *Nature's* columns. On page 473, at the end of the list of acknowledgements in a neuroscience paper, is the following: "R.R. conceived the experiment, and together with A.H. and L.L. carried it out; C.D.B. designed and carried out the data analysis; R.R. and C.D.B. co-wrote the paper." *Nature* decided to accept the authors' request to publish this form of words — in the light also of previous concerns about transparency in relation to scientific misconduct, as well as proposals from correspondents (see, for example, page 405). This policy of allowing authors to succinctly describe their contributions now applies to anyone who requests it.

This is part of a movement that, we hope, will spread naturally across the scientific community, it having already become estab-

lished in some biomedical journals. Experience in the pages of *The Lancet* has recently been analysed by Veronica Yank and Drummond Rennie. Their report is at <http://www.cbe.org/cbe/Yankrev.html>. A proposal of standards for such listings and other useful references can be found at <http://www.cbe.org/cbe>.

This policy is experimental. We believe that, for now at least, authors and editors are capable of jointly deciding what works best in particular circumstances. If, for instance, a long list proves desirable, we may consider it for the web version of *Nature* only.

We hope that, as the practice spreads, the dishonourable practice of "honorary authorship" — authorship by virtue only of seniority, for example — will diminish. More positively, we hope it will lead to a fuller appreciation of just who made what critical contribution. □



German research bodies urged to open up more to new ideas

MPC/FILSER

[MUNICH] An external evaluation committee has urged Germany's two main basic research organizations to stop living on past glories and to embark on reforms that will make them more responsive to modern research needs. Federal and regional governments, says the committee, should help by relaxing restrictive employment laws.

The committee was set up in 1997 by the Bund-Länder Kommission (BLK) — the body that coordinates regional and federal research policies — to assess the procedures of the Max Planck Society (MPS) and the Deutsche Forschungsgemeinschaft (DFG), Germany's university funding agency.

Ten non-German scientists and science-policy experts were asked to consider whether the bodies' organizations allowed them to support the best research and respond quickly to new research areas.

The committee's report, submitted to the BLK last week, includes few surprises. The two research organizations, however, say it will provide moral support for important reforms that are already partly under way. But they stress that not all the recommendations



MPS's Markl, left, and DFG's Winnacker: keen to improve the lot of Germany's young researchers.

are likely to be implemented; only those compatible with their fundamental principles — in particular, their political autonomy — will see the light of day.

The report highlights many well-known problems in German science, including inadequate academic career opportunities for young researchers, restrictive employment laws that prevent organizations from offering competitive salaries to top

researchers, and working conditions that hamper the progress of women.

A general problem in both universities and MPS institutes is the time and effort needed to establish a research career. The report says that *Habilitation* — the German postdoctoral teaching and research qualification required by most universities for faculty membership — should be abandoned, because it delays careers unnecessarily. More funding, it says, should be available for junior independent research groups.

Such comments are music to the ears of Ernst-Ludwig Winnacker and Hubert Markl, respective presidents of the DFG and MPS. Both are champions of young scientists, and both have created programmes to support independent young-scientist research groups in the past few years.

But other comments in the report are more difficult for them to swallow. The DFG was criticized for its "tendency towards a conservative approach" to science funding. The report says the agency should "not only respond to long-term developments, but should actively influence them", acknowledging the DFG's success in funding mainstream research, but criticizing its record in funding interdisciplinary research and work in new areas.

The report also says that the DFG's peer-review system needs to be more responsive to new ideas. In response, Winnacker says he is aware that the DFG needs to break out of its self-reproducing peer-review system, and says mechanisms for this are in place.

The DFG's core of about 500 referees is elected by German academics from the nominations of the country's scientific societies, and organized into 37 committees (*Fachausschüsse*) divided according to discipline. This core, which is dominated by older, more established scientists, is supplemented by three times as many outside referees.

These outside referees may be called upon more often in the future, says Winnacker, saying that "in the past few years we have managed to reduce the average age of the members of our selection committees by ten years". But he adds that there are limits to how far DFG staff can go in selecting referees.

The DFG prides itself on being an organization run by and for researchers. Many scientists believe that most of its decisions should only be made by representatives elected by the community. "We will need to discuss the issue in depth," says Winnacker.

Winnacker is also uncomfortable with the evaluation committee's call for the DFG to develop 'strategic' research programmes,

US labs braced for anti-spying legislation

[WASHINGTON] The US Congress is expected to pass legislation in the near future restricting foreign scientists' visits to nuclear-weapons laboratories. The move is a response to accusations of spying by China, which reached their peak last week with the release of a report by a Congressional committee chaired by Christopher Cox (Republican, California).

Scientific societies are working to influence the legislation, which is certain to be passed in response to the scandal. Despite public expectations of a clampdown on partnerships with foreign scientists, laboratory officials and the societies are confident that the legislation will be framed so as to allow partnerships to continue.

Critics of the Clinton administration in Congress are talking about clamping down on foreign visits to the laboratories, but in practice their room for manoeuvre is constrained by the need to keep the weapons laboratories scientifically competitive.

"The Congress is attempting a legislative fix to a problem that isn't amenable to a legislative solution," says an official at one scientific society. "A large number of people, including foreign scientists, come into occasional contact with people who do the classified work. There is no way to stop that

without harming the US nuclear weapons programme."

Defence bills passed by the Senate, and expected to be passed by the House of Representatives after the recess, include special provisions on laboratory security. The laws will impose conditions on visits by scientists from so-called 'sensitive countries' to the three nuclear weapons laboratories: Lawrence Livermore in California, and Sandia and Los Alamos in New Mexico.

A proposal from Jim Ryun (Republican, Kansas) would drastically curtail foreign visits to weapons laboratories. But a more pragmatic package proposed by Norm Dicks (Democrat, Washington), the senior Democrat on the investigative committee, and backed by Cox, is likely to form the basis of the final legislation, observers say.

Under this proposal, visitors to the laboratories from the sensitive countries, which include Russia, China and India, would face a 60-day moratorium until new vetting procedures have been implemented.

The parts of non-weapons laboratories which these scientists normally visit will not be affected. But some of these, including Brookhaven, have already stepped up their vetting procedures in response to the Chinese spying scandal.

Colin Macilwain

QUERBACH

making it more than an organization that simply reacts to ideas from the community. He argues that a 'top-down' role for the DFG may be incompatible with its autonomy, a characteristic much acclaimed in the report.

Although the DFG does not oppose the principle of creating new programmes, Winnacker says "it is important that these are selected either by the research community itself, or in an interplay between DFG staff and the research community".

The report recognizes that the work of the DFG has grown enormously over the past decade, increasing the burden on DFG staff and their referees, and it recommends that the number of elected referees be increased by a quarter.

This suggestion, which is already being implemented by the DFG, is particularly welcome, says Klaus-Peter Hoffmann, professor of zoology and neurobiology at the University of Bochum, and a biology referee for the DFG for 25 years. Hoffmann receives 300 grants to review every year, and says the pressure crushes creativity. "When overburdened, one tends to shy away from risk, and this is one reason we have tended to avoid entering new territory," he says.

The MPS, while praised for the quality of its research, comes in for criticism for being "isolated from the university system" and for sticking too closely to the 'Harnack princi-

ple' on which it is based. This principle causes research to be concentrated in institutes headed by directors selected for their scientific stature rather than their research area.

Directors are appointed for life, receive generous support for their research and have complete freedom to choose the direction of their research. But the report says that this could hamstring the society by tying up too many resources in a single person for decades.

The report also calls for "International Max Planck Research Schools" to be created at universities to allow graduates to benefit from the expertise of local Max Planck institutes. It says that time-limited Max Planck research units should be created at universities. Both measures, it says, would help bring the MPS and universities into closer contact.

Although Markl supports cooperation with universities, he questions whether the committee, who did not review university research, was aware of the close cooperation that already exists in some places. He points out, for example, that "ten per cent of MPS directors now have university chairs".

But Axel Ullrich, a director of the Max Planck Institute for Biochemistry, near Munich, says that the current links are patchy and dependent on personal contacts. Ullrich has not been made a university professor, because most of his career was spent in the

United States, "so I was not established in academic circles in Munich". He welcomes closer contact between the two institutions.

Markl is in favour of international research schools. But he is "not yet sure" of the value of creating Max Planck groups in universities, as this moves away from the Harnack principle and, he says, could stretch resources too thinly.

Markl says that reforms will have to be paid for out of the five per cent annual budget increase promised by the German gov-



Brook: backs 'top-down' guidance.

ernment for the next few years. "My first priority is to complete our plans for building up MPS activities in East Germany," he says. Future reforms will depend on whether the MPS can afford them.

Richard Brook, head of Britain's Engineering and Physical Sciences Research Council and chair of the committee, says that the MPS and DFG need to become more accessible to young scientists and women — "we were shocked to realize that only three per cent of MPS directors are women". Brook also says that the DFG should not necessarily be afraid of giving 'top-down' guidance. **Alison Abbott & Quirin Schiermeier**

RAL PHOTOSERVICES

Proceed with caution, says UK report on ethics of GM foods

[LONDON] Last month's decision by the British government to strengthen the monitoring of genetically modified (GM) foods but resist calls for a moratorium on their commercial planting has been implicitly endorsed by the main UK bioethics advisory body (see *Nature* 399, 287; 1999).

In a report published last week, the Nuffield Council on Bioethics says that there are many aspects of the development and introduction of GM foods that warrant firm government action (see page 405 in this issue). These range from steps to limit the breadth of patent claims to what it describes as a "moral imperative" to develop GM staple foods for the Third World.

But the report dismisses opposition to GM crops based either on broad claims about their 'unnaturalness' or on their potential for misuse. The genetic modification of crop plants, it says, "does not differ to such an extent from conventional plant breeding or other human interventions with the natural world as to make the process morally objectionable in itself".

The overall message has been welcomed by the agrobiotech industry and the government. Jack Cunningham, the cabinet minister responsible for the government initiatives, said the report was "independent backing for the government's approach".

The response has been cooler from environmentalist, Third World and religious groups, upset that the report's admittedly utilitarian stance does not pay closer heed to the ways that GM technology has been — and is likely to be — used.

The report, *Genetically Modified Crops: The Ethical and Social Issues*, was drawn up by a working party chaired by Alan Ryan, professor of philosophy at the University of Oxford. It outlines three areas in which, it claims, moral considerations are relevant: general welfare, people's rights (for example, to freedom of choice as consumers) and the principle of justice — the fair sharing of burdens and benefits.

From this perspective, while endorsing the patenting of genetic sequences, it urges national and international patent bodies to discourage patents "which allow extensive control over a single crop species," and to draw up new guidelines for doing so.

The report also strongly endorses the use of GM crops in the Third World, arguing that the British government should allocate a "substantial amount" of its increased aid budget to research and development on GM food staples grown in developing countries.

But it also urges careful attention to the potential environmental impacts of GM crops, as well as to ensuring that farmers in

these countries are given a choice between GM crops and traditional varieties.

By sticking to its three ethical principles, the working group rejects the demands of some critics to take a broader ethical stance. "It is the *deleterious consequences* of our farming techniques to our environment and public health, not their 'unnatural' character that should preoccupy us," the working group says.

Thus, although supporting in principle the labelling of GM foods, it opposes labelling as "not necessary or practical" for foods produced by a GM process where no difference can be detected.

The Third World group Christian Aid, which recently issued a report arguing that GM crops were unnecessary for problems that could be resolved by better food distribution (see *Nature* 399, 99; 1999), claimed that the council was "out of touch" with the real causes of hunger.

Donald Bruce, who heads a project on GM foods for the Church of Scotland, says that although he agrees with much of the working party's conclusions, the report's dismissal of the 'unnaturalist' position is "a dogma rather than a serious argument".

But Ryan defends the working party, saying that the report "is intended to be read in a highly pluralistic society". **David Dickson**

US Senate gets tough on animal activists

[WASHINGTON] Following attacks by animal-rights activists on research facilities in Minneapolis and San Francisco, the US Senate has moved to toughen federal penalties for such crimes — including the possibility of the death penalty if anyone is killed in an attack.

The new measure also requires the Federal Bureau of Investigation (FBI) to establish a 'National Animal Terrorism and Ecoterrorism Incident Clearinghouse'. This would centralize information on attacks gathered by federal, state and local law enforcement agencies, which have had difficulty making arrests and convictions in cases of lab vandalism.

An amendment aimed at those convicted of attacking animal facilities — whether zoos, fur farms or research labs — and causing more than \$10,000 worth of damage was included in a broad juvenile crime bill that was passed by the Senate late last month.

Under the amendment, offenders would face up to five years in prison and fines of up to double the cost of the damage caused. If explosives or arson were used, prison terms would be at least five years and up to 20.

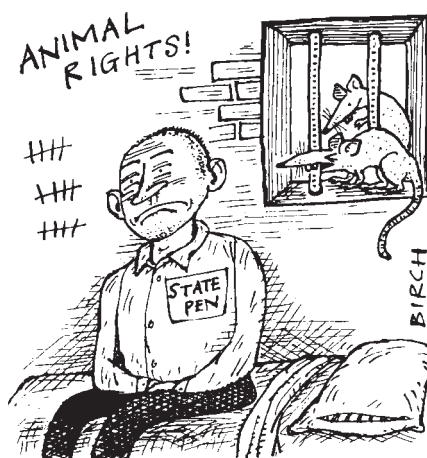
The amendment was written by Senators Orrin Hatch (Republican, Utah) and Dianne Feinstein (Democrat, California), and passed by a vote of 85 to 13. It toughens the criminal penalties in the Animal Enterprise Protection Act of 1992, which dictated a maximum prison term of one year for attacks causing property damage, and left unspecified the prison term for attacks in which anyone is killed.

"[The amendment] is a useful move," says Walter Low, a professor of neurosurgery at the University of Minnesota whose lab was ransacked by activists in April (see *Nature* 398, 553; 1999). The Animal Liberation Front claimed responsibility. "Hopefully it will cause these people to think more carefully about their actions. This kind of violence and vandalism has no place in a civilized society."

Research lobbyists also praised the measure. "It gives the Animal Enterprise Act more teeth," says Mishka McCowan, programme director at Americans for Medical Progress, a group based in Virginia that was set up to counter the animal-rights movement.

McCowan says the vandalism at the University of Minnesota — and a smaller-scale attack less than three weeks later at the University of California at San Francisco (see right) — were prompting "concern that this would become a trend. The stiffening of the penalties alleviates that a little bit."

But animal-rights activists say the amendment will not make any difference. "Nobody in any social movement has ever been deterred from breaking and entering or arson because the penalties have been elevated," says Ingrid Newkirk, president of People for



the Ethical Treatment of Animals (PETA), based in Norfolk, Virginia. She calls the amendment "a knee-jerk piece of legislation".

Newkirk says PETA uses only peaceful civil disobedience, but adds that researchers have brought the more violent incidents on themselves. "The laboratories, because they fight even the most modest proposals to improve the living and dying conditions for animals in their enterprises, are responsible for an escalation of [violence] by people frustrated that they can't achieve change through normal channels."

Civil libertarians challenged what they called the amendment's unconstitutional focus on animal-rights activists — especially when invoking the death penalty. "We think

that criminal laws should be equitably enforced across the entire population," says Rachel King, legislative counsel at the American Civil Liberties Union. "No one group should be singled out for harsher punishment than any other."

King claims that the death penalty provision was introduced surreptitiously into the bill. Most senators, who were voting on nearly 200 amendments to the bill, had no idea that the animals amendment contained a new capital crime, she says.

The amendment, which contains many other provisions dealing with youth gangs, originated with Hatch, a member of the advisory council of Americans for Medical Progress. In his home state, a youth gang that eschews the eating and wearing of animal products has attacked animal facilities.

Barbara Rich, executive vice-president of the National Foundation for Biomedical Research, says the FBI clearinghouse is particularly welcome. "We've had very few successful prosecutions" after attacks on research labs, she says. "Anything that will encourage local and federal law enforcement to pursue these things vigorously we think is a good idea."

The House of Representatives is expected to take up a juvenile crime bill later this month. If this does not include a similar provision, the measure could still be retained in a final version negotiated in a House-Senate conference.

Meredith Wadman

San Francisco scientists hit by growing protests

[WASHINGTON] When a dozen laboratories at the University of Minnesota's Twin Cities campus were attacked in early April, word of some \$2 million worth of damage and 116 lost animals spread quickly in the scientific community.

Less attention was given to an attack on three labs at the University of California, San Francisco (UCSF), on 23 April. In conjunction with a peaceful protest by about 60 demonstrators, several labs were broken into by a smaller group.

This group threw computers and centrifuges on the floor in the Cell Culture Facility, and damaged surgical equipment in a cardiology lab. Activists then entered a neurology lab and released four transgenic

mice — two with scrapie — being used to develop diagnostic methods for 'mad cow disease'. Three activists were charged with criminal trespass. The damage is estimated to cost \$11,000.

The incident is part of an unsettling pattern, says Ara Tahmassian, UCSF's assistant vice-chancellor for research services. Peaceful demonstrations have been occurring regularly on the campus for a year, he says. But three times in the month before the campus break-in, activists demonstrated at the home of Steven Cheung, an otolaryngologist who studies hearing loss using squirrel monkeys as a model. What Cheung believed to be a burning effigy of himself was thrown at the house.

"There is a kind of radicalism that we haven't seen before. I think we [in the research community] need to be a lot more careful not to take this lightly," says Tahmassian. "My message to the campus has been: be on guard, times have changed."

In Minnesota, police have made no arrests. But Kevin Kjonas, the Animal Liberation Front spokesman who said the group claimed responsibility for the attack, has been subpoenaed to appear before a federal grand jury this month. Police will not say whether he is a suspect in the attack.

The Minnesota legislature passed a bill last month imposing civil damages of at least \$5,000 for the unauthorized release of laboratory animals.

M.W.

'Frontiers' grants count cost of success

[MUNICH] Ten years after a shaky start, the Human Frontiers Science Program (HFSP) has established itself as one of the leading intercontinental research grant programmes, and as an increasingly important source of funds for collaborative research in molecular biology and neuroscience.

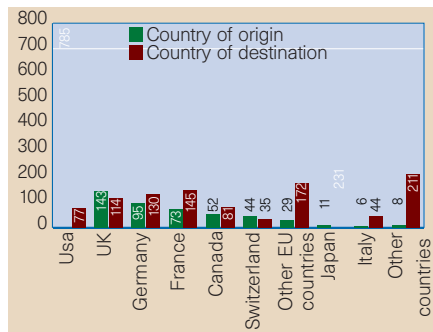
But, as the programme celebrates its birthday, the HFSP board is struggling with its future financing. (The first celebration took place last December in Tokyo, and the final one will be in December in Washington — with a ceremony and two scientific workshops in Strasbourg, France, this week.)

At present, only one in eight applications for research grants and one in four applications for long-term fellowships can be approved for funding. Some of the best are refused. "We can only afford to support around half of the applications which our referees tell us are very high priority," says Piergiorgio Strata, professor of neurosciences at the University of Turin and a member of the HFSP board of trustees.

Akiyoshi Wada, director of the RIKEN Genomic Sciences Centre and one of the principal architects of HFSP, confirms that "the situation will have to change to be able to continue funding promising areas of biological research".

Human Frontiers was launched at the initiative of the Japanese government in 1989, and has distributed US\$414 million to researchers around the world. Initially, however, it was widely distrusted for being motivated as much by political as scientific considerations.

The feeling was particularly strong among research groups in Europe and the United States who felt that Japan — which had a relatively poor track record in basic research — had set up the programme to exploit their



HFSP long-term fellows: where they come from and where they go.

work. But this attitude has been "completely reversed" after ten years of genuine support for international research, says Wada.

"There were several misunderstandings in the past, which led people to think that HFSP was an attempt by Japan to promote its technological development," says Wada. "Fortunately the concept of the programme — to support collaborative basic research through a common funding mechanism — has now been fully understood."

By the end of the month, the Human Frontiers board is expected to select a successor to Michel Cuénod, who steps down as secretary general next spring. There are three candidates, all internationally renowned scientists. One is Torsten Wiesel (75), Nobel laureate in 1981 for work on the neurophysiology of vision, and currently president emeritus of the Rockefeller University in New York.

The second is Albert Aguayo (65), a professor of neuroscience at McGill University in Montreal, Canada. Aguayo, a native of Argentina, works on neuron regeneration and is a former chairman of the Human Frontiers council of scientists. The third candidate is Lennart Philipson (70), a molecular biologist at the Karolinska Institute in Stockholm who formerly headed the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany.

The programme has become one of biology's best-regarded grants schemes. But its success has led to oversubscription, and the board remains keen to increase the budget to raise the success rate for applicants.

One of the main tasks facing the next secretary general will at the same time be to redress the continuing funding imbalance between Japan and its partners. Japan provided nearly all the funds for the first three years of the programme and has contributed 80 per cent since.

But the scientific benefits reaped by Japan remain lower than those of other partners. Within the research grants section, for example, which absorbs a third of the budget, 32 per cent of investigators have come from the

United States, and only 16 per cent from Japan. Yet the United States will contribute less than 10 per cent to the budget next year.

A new basis for funding was proposed in 1997. Partner states — apart from Japan — agreed in principle to increase their contributions over five years, up to a level related to their gross national product (see *Nature* 387, 446; 1997). Japan's contribution would remain constant until it fell to 50 per cent of the total.

The new formula would increase the annual budget from \$45 million in 1997 to \$60 million in 2002. But few countries are on target to meet the non-binding increases. Already the 1999 budget of \$47 million falls well short of the promised \$53 million.

The United States wins out in the long-term fellowship programme, as well as in the research grants programme (see left). While the United States remains a Mecca for young scientists from other countries, its own researchers are less convinced of the benefits of doing a postdoc abroad.

The stature of the United States gives its scientists little incentive to leave, says James Sutherland, a mouse geneticist working as a Human Frontiers fellow at the EMBL outstation at the Adriano Buzzati-Traverso campus at Monterotondo near Rome.

Sean Oldham, a developmental geneticist from North Carolina, now working at the University of Zurich, says he applied for a Human Frontiers fellowship because of its prestige value — but most colleagues told him it would be a "retrograde career step to go to Europe".

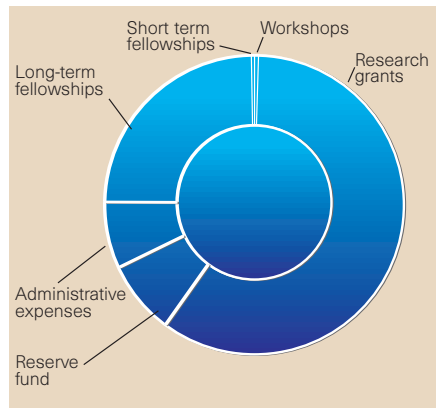
Oldham says his colleagues felt he would find it harder to get a job in the United States, because of loss of contact with US research networks, and because a European postdoc would not be taken seriously. Scientists, he was told, "do not work hard in Europe".

But Oldham says that the stereotype of Europe as a scientific playground is false. Both he and Sutherland believe that, provided they maintain contacts with US scientists, they will have no trouble returning to the US academic system.

"Working in a foreign laboratory broadens your horizons — your personal human frontiers — enormously," says Oldham. It is valuable to learn that there are other successful research cultures apart from the United States, he says.

For example, he has been impressed by "the extraordinary meticulousness" of Swiss researchers, but he has been taken aback by "the extraordinary respect that Swiss scientists maintain for the authority of the professor". Overall it is different, and it works, he says: "After all, Switzerland has the highest density of Nobel prizewinners in the world."

Alison Abbott & Asako Saegusa



The HFSP has spent a cumulative budget of US\$414 million since its launch in 1989.

Research grants average US\$230,000, and teams receiving them must have an international make-up. Eighty per cent of the 422 awarded so far involve scientists from at least two continents, and a third involve three continents.

Californian grad students vote to unionize

[SAN DIEGO] Graduate students at the Berkeley and Los Angeles campuses of the University of California have voted overwhelmingly for union representation as part of a statewide bid for better pay, benefits and working conditions for teaching assistants.

Elections at the remaining six of the university's campuses are due to be completed next week. A state agency will then count the votes to determine whether other campuses will also be represented by affiliates of the United Auto Workers Union.

The elections follow a 16-year campaign by graduate students at the university to secure union recognition for teaching assistants, who have increasingly played a vital role in academic education, in science, engineering and mathematics. They typically manage laboratory sessions, grade papers and staff offices for professors, who limit their involvement to giving lectures.

During this period, California's two Republican governors contested unionization at every opportunity. But the recent election of Gray Davis, a Democrat supported by unions, whose position gives him direct involvement in the administration of the University of California system, is seen as prompting a change in philosophy.

There has long been discontent among teaching assistants at the university, who complain of excessive workloads and a lack of benefits, such as adequate health insurance and dental and optical cover.

The campus elections have been allowed to proceed in recent weeks when the university's attorneys stopped fighting attempts to organize them, after spending about \$1.7 million on legal fees in the past four years alone to block teaching-assistant unions.

Similar concerns by teaching assistants have emerged at other campuses around the country. There are unions representing graduate students at some 20 institutions.

Teaching assistants at New York University last month filed a petition with the National Labor Relations Board to hold an election for representation by the United Auto Workers Union.

This petition is believed to be the first such request for union representation by graduate students at a private university. For the past 25 years, the National Labor Relations Board has held that graduate students who perform teaching and research duties are primarily students, and so do not have collective bargaining rights under the National Labor Relations Act.

At New York University, housing and fee reductions are a major issue of concern for the teaching assistants, says Sivan Rottenstreich, who is in his third year of a mathematics doctoral programme at the university's Courant Institute. "The big issue is to



Campus protest: teaching assistants complain of being overworked and lacking health benefits.

have a voice; we can improve this university."

University of California graduate students from San Diego to Berkeley voice widespread displeasure with the workload and lack of input in the workplace. For example, Jenny N. Hyatt, a first-year doctoral student in materials science at San Diego, says that as a teaching assistant she was assigned at times to eight three-hour lab sessions a week, but was being paid to work only 20 hours. "I was

being overworked," says Hyatt. "When I complained, they ignored me."

Most teaching assistants say that faculty members largely remained estranged from the graduate-student organizing effort, with some supportive and some opposed.

The University of California administration has "stalled for a long time in the face of evident student demand for a union," says Michael Watts, director of the Institute for International Studies at Berkeley, who was involved in organizing assistants as a graduate student at the University of Michigan. "There is a need to have a systematic way to regulate workloads, pay and benefits."

Teaching assistants voted by 833 to 293 to unionize at Berkeley. At Los Angeles, there was a similar turnout in March among the 1,700 teaching assistants, who voted by 718 to 269 to unionize. Voting took place last month at San Diego, Irvine, Davis and Santa Cruz, and teaching assistants at Riverside and Santa Barbara will vote in the next week.

Contract negotiations are to begin this summer between the University of California and student groups that vote for a union. There is "absolutely genuine sentiment" to bargain with the teaching assistants, says Joseph Duggan, associate dean of the graduate division at Berkeley.

Rex Dalton

Reprieve for neutron project in Congress

[WASHINGTON] The Science Committee of the US House of Representatives has agreed to support funding to begin building the Advanced Spallation Neutron Source at Oak Ridge in Tennessee, relieving fears that the project would fail to win money from Congress.

The committee voted unanimously last week to allocate \$100 million to construction of the project, provided that the Department of Energy meets a list of conditions. Although the money allocated by the committee for the fiscal year 2000 is barely half the amount requested by the Clinton administration in February, the vote was interpreted as a reprieve for the project.

James Sensenbrenner (Wisconsin, Republican), chair of the committee, had previously wanted to withhold funds altogether (see *Nature* 398, 739; 1999).

On the same day last week, the Senate energy and water appropriations subcommittee, chaired by Pete Domenici (Republican, New Mexico), voted to provide \$169 million for construction of the project, serving notice that it still has powerful support in Congress. Los Alamos laboratory in New Mexico will build the accelerator that will drive the neutron source.

The Science Committee's compromise

was struck after a heated debate on a Democrat amendment to restore \$150 million to the project, during which it became clear that several Republicans were unhappy with Sensenbrenner's plan to withhold funding.

Republicans Connie Morella (Maryland), Judy Biggert (Illinois) and Vernon Ehlers (Michigan) each spoke for the project, but voted against the amendment to restore funding. Two moderate Republicans, Sherwood Boehlert (New York) and Roscoe Bartlett (Maryland), were absent for the vote, and a third, Joe Barton (Texas), voted with the Democrats.

The result was a vote tied at 17-17, causing the amendment to fail. But it was obvious that the committee did not want to pass a bill with no funding for the project, which has wide support in the scientific community. When the committee broke for lunch, staff thrashed out a compromise that would give the project \$100 million.

Congressional staff say that the appropriations committees, which have the real budgetary power over the project, are likely to agree to provide at least \$150 million next year, but will attach conditions to ensure proper management by the energy department.

Colin Macilwain

Spanish council quizzed on recruits...

[BARCELONA] Spain's Higher Council of Scientific Research (CSIC), the country's main public source of science funding, has created 150 new positions for junior scientists this year, to be distributed between more than 100 institutes nationwide.

But the optimal allocation of these jobs is still a hot topic, particularly following the recent debate on personnel selection at universities, where nepotism is said to be rife (see *Nature* 396, 712; 1998).

In particular, there are frequent complaints about 'local lobbies' at CSIC institutes, which draw up job profiles that favour predetermined candidates (see box).

Both César Nombela, the CSIC's president, and Emilio Lora-Tamayo, vice-president of scientific and technological research at the CSIC, are enthusiastic about the substantial increase in new posts, promised by the government 18 months ago, to assist young researchers returning from abroad.

In 1998, 120 new positions for junior scientists were created, of which 65 were tenured staff positions and 55 non-permanent (interim). This year's employment offer, which

opened in March, will allow the incorporation of still more junior scientists, creating 60 tenured and 90 interim positions.

Mariano Esteban, head of the CSIC National Centre of Biotechnology in Madrid, is confident that the CSIC's appointment procedures will ensure that those appointed are of a high standard. He says that periodic evaluation of CSIC centres by committees of international experts should encourage competitiveness in research and make researchers more mobile.

Lora-Tamayo says the jobs requested by each institute do not determine the final allocation of positions. Potential positions are first analysed by 'coordinators' in eight scientific areas, and finally decided by an appointment board.

Although each institute is asked to prioritize its needs for new appointments, he points out that "this is only a consultative step", and that the CSIC Presidency aims "to make job descriptions as open as possible to offer the best possibilities to the better applicants".

The needs of each institute are then evaluated, taking into account its scientific output



Grand designs: CSIC's task is to distribute the new posts between 100 institutes nationwide.

and ability to raise external research funds, says Lora-Tamayo. The selection process, he says, is designed to ensure the recognition of equal opportunities and scientific merit.

Luis Sanz-Menéndez, of the CSIC Institute for Advanced Social Studies, says that, in contrast to university appointment panels, which can set their own evaluation criteria, "CSIC panels must follow general criteria laid down by the agency". He admits that problems occur at some institutes, but says that these may reflect organizational problems, for example the lack of a collective strategy.

Alfonso Vázquez-Vaamonde, professor of chemistry at the CSIC National Centre of Metallurgical Research and president of the CSIC's association of research personnel, acknowledges that favouritism can operate in the appointment process, mainly when there is a local candidate who may not have got a post previously.

Nombela accepts that the CSIC's appointments policies could be improved. But he says that "the analysis of the institutes' proposals is becoming increasingly rigorous", and points out that recent indicators published by the European Commission lists the CSIC as the only Spanish institution in the first 15 on the publications-impact index.

The issue has been highlighted by the difficulties encountered by many postdoctoral students returning from abroad at the government's invitation.

Fernando Valladares, an ecologist at the CSIC Centre of Environmental Sciences, says that, "although new posts have been created, general deficiencies continue". In particular, he calls for "new forms of recruitment to allow higher levels of dynamism".

Marta Alvarez Fernández, a chemist who last year won a permanent research position after applying for the CSIC public employment offer, is critical of what she describes as "improvisation in scientific policy".

Manuel Torres, a staff scientist at the CSIC Institute of Applied Physics, confirms the criticisms, although he argues that the problems appear to be worse at institutions doing applied research than at CSIC centres devoted to basic research.

Xavier Bosch

... as physicist's complaints hit their mark

[BARCELONA] A leading physics research institute in Spain has agreed to broaden its recruitment criteria for three new posts, following a protest by one of its professors that its criteria were too narrow, and had been written to favour particular candidates.

Javier Bermejo, a professor of physics at the Institute of Structure of Matter in Madrid, says that junior staff recruited to his institute last year – a high-energy theorist, a nuclear physicist and a vibrational spectroscopist – were appointed through open and fair competition.

But he complained that, this year, his institute's recruitment committee had again identified its top priorities as being specialists in the same three fields. According to Bermejo, although institute officials claimed that "balanced development" was needed, the narrow designations reflected pressure to appoint unsuccessful applicants from

the previous year.

Francisco José Balta-Calleja, the institute's director, says there were inevitably internal conflicts because "each department may seek a post for its own candidate".

According to José Vicente García-Ramos, head of the department of vibrational spectroscopy and a member of this year's committee, one problem is that there are scientists with excellent research records in each department, either on short-term contracts or temporarily working abroad, who cannot be appointed because of insufficient posts.

Bermejo says that, in general, he applauds the CSIC's efforts to increase the number of scientific staff and diversify research in its centres (see main story). But he is concerned that local pressures contribute to "in-breeding" that may close doors to scientists trained in emerging research areas.

Such pressures, he adds, tend to come from researchers who are active in

fields that are already overstaffed, but who have the power to maintain a "self-replication" process.

"My own career path to the position of professor was made possible by open competition," says Bermejo. "Young researchers should have to apply for posts that are as general as possible, and not narrowly defined."

The CSIC says it has long been aware of this issue. After indicating that he intended to go public with his concerns, Bermejo says he was told by Emilio Lora-Tamayo, vice-president of scientific and technological research at the CSIC, that the priorities identified by his institute's committee were too narrow, and that the CSIC accepted that this year's posts at his institute should include broader profiles.

The list of profiles was announced last week, and the posts in question are broadly defined as theoretical and experimental physics – without additional designations.

X.B.

Canadian institute seeks to secure its financial future

The Canadian Institute for Advanced Research promotes excellence in research. But as a funding agreement with the government is set to expire, it must secure a firm financial base if this success is to continue.

Colin Macilwain

[BANFF, CANADA] The Canadian Institute for Advanced Research (CIAR) last week held the second Congress of its 200 fellows and associates from Canada and around the world.

But Fraser Mustard, the fiery former physician and medical researcher who founded CIAR 17 years ago, despite hearing how the institute has promoted outstanding scientific progress, pronounced his own role as having been a failure in one important way.

"I've failed because we couldn't develop an endowment basis for the institute," Mustard told the meeting, reflecting on the perennial struggle that this "institute without walls" has faced in raising money to cover its operating costs. "The institute has to have an endowment to have any chance of surviving over the long haul," says Mustard.

The Canadian government is now considering a proposal from CIAR under which it would match the money the institute is able to raise from private sources and from Canada's provincial governments.

Research programmes

Kevin Lynch, deputy minister of Industry Canada, says it is too early to say how this proposal will be received. But, given the institute's research track record and the leverage it has exercised by distributing a little money to key people in multidisciplinary, geographically scattered groups, it has a reasonable chance of being accepted.

CIAR pays a small amount of money to the universities that employ its fellows and associates in exchange for a lightening of their teaching load. This has enabled them to devote more time to research, and especially to collaborate with the other members of their programme. They also receive money to enable them to meet regularly.

The approach is well suited to Canada's circumstances. Undergraduate teaching loads are heavy in Canadian universities, and excellence in the universities is widely distributed. The establishment of 'centres of excellence' tends to be politically divisive. And the country's small scientific elite knows it must work closely with scientists overseas,



Dupré: under pressure to broaden institute's role.

including the large Canadian diaspora in Europe and the United States.

The approach has provided a substantial boost to most of the eight programme areas that CIAR supports. Paul Hoffman, for example, a geologist formerly with the Geological Survey of Canada and now at Harvard University,

says that CIAR's Earth system evolution programme "was extremely important" in expanding his work in plate tectonics to produce his celebrated 'snowball Earth' thesis, which holds that the Earth suffered a series of violent ice ages 700 million years ago.

Franz Lang, a biochemist at the University of Montreal, says his fellowship in CIAR's evolutionary biology programme helped to propel him into the elite circle who publish in top journals and receive substantial research grants. The programme allows them "to compete with the most prestigious and best-funded American universities," says the programme's director, Ford Doolittle of Dalhousie University, Nova Scotia.

Some of the research programmes merge social sciences with hard sciences, and reflect a Canadian perspective on social issues. The programme on population health, for example, has charted the importance of social factors as markers for disease, and shows that Canadian social cohesion does more for public health than the far larger per-capita expenditure of the United States on medicine.

CIAR has succeeded in establishing pre-eminent, internationally networked groups, but it has struggled to secure a firm funding base. It relies on wealthy individuals and provincial governments — chiefly that of Ontario — for its operating funds.

Under an agreement that expires next March, the Canadian government provides

Can\$1 (US\$0.68) for every Can\$2 raised outside, bringing CIAR's annual income to around Can\$10 million. Financial pressure almost closed it in 1994, and it could face the same situation when the private donations that came to its rescue expire.

The institute is also ageing, leading to tensions between old programmes and new ideas. Some scientists with long-standing links with CIAR argue it is in danger of losing its way under the direction of Stefan Dupré, a political scientist from the University of Toronto, who replaced Mustard in 1996.

They fear that it will embark on 'sexy', problem-solving exercises in applied science, and lose sight of its goal of supporting the best people in their quest for basic knowledge. The 30-strong CIAR research council is split on the question, according to one of them.

A university without walls

Carolyn Tuohy, the deputy provost of the University of Toronto, who recently led a study into how CIAR should deal with its established programmes, alluded to these tensions in an open discussion on CIAR's future: "Are we becoming a university without walls, in more ways than one?" she asked.

CIAR programmes are supported for five-year periods, and some participants say that groups achieve most during the second or even third five-year period. But Howard Alper, vice-rector of research at the University of Ottawa and a member of the CIAR research council, asserts that "it is crucial that sunset clauses be put in and carried out".

The research council has rejected that idea, says Dupré. Officials hope that some programmes will evolve into self-sustaining research centres with other sources of support. It may also phase out the salary support on its mature programmes, while continuing to pay for meetings and travel, enabling more new programmes to be established.

CIAR may also take steps to address what Lorna Marsden, the president of York University, Ontario, bluntly termed its "white, male, club-like atmosphere". But she hastily added that this had to be done while maintaining "the naked elitism that is CIAR".

Participants in the congress called on Dupré to raise CIAR's public profile and to broaden its role in education and in solving problems of public policy, such as those surrounding genetically modified foods. Tom Brzustowski, president of the Natural Sciences and Engineering Research Council of Canada, said CIAR could seek to advise the government on scientific matters. "There is a great deal that this programme has already taught the rest of research in Canada," he says.

But most participants seemed to agree with Bill Unruh of the University of British Columbia, a former director of the CIAR cosmology and gravity programme. To powerful applause, he called for CIAR to stick to its strengths — and support basic research. □

Proposal to double US spend on information technology research

[WASHINGTON] The Science Committee in the US House of Representatives plans to introduce legislation to double spending on basic research in information technology over the next five years.

The proposal, due to be announced this week by committee chair James Sensenbrenner (Republican, Wisconsin) in a speech to Silicon Valley executives in Palo Alto, California, would give \$4.8 billion to computer research across six agencies, including the space agency NASA and the National Science Foundation.

Sensenbrenner does not support a Senate proposal to double all research spending. But support for the information technology measure may be intended to show that the fiscally conservative legislator supports more discriminating increases in areas of research that he thinks show particular promise.

Call for ban on Japan's 'scientific whaling'

[TOKYO] Member states of the International Whaling Commission last week called for a ban on Japan's 'scientific whaling' at their

annual meeting on the Caribbean island of Grenada. Dismissing Japanese claims that the catches are to investigate the level of whale stocks, the commission voted for a non-binding resolution to ban such whaling.

The meeting also rejected Japan's proposal to catch 50 minke whales in the Antarctic. Japan caught 500 whales last year, and has been criticized for carrying out what is believed to be a commercial operation in disguise, by selling whale meat after the completion of its 'scientific research'.

Republican hits out at stem-cell ethics report

[WASHINGTON] A prominent Republican member of the House of Representatives has attacked last month's finding by President Bill Clinton's National Bioethics Advisory Commission that the government should fund embryonic stem-cell research and the extraction of such cells (see *Nature* 399, 292; 1999).

Tom Bliley (Republican, Virginia), chair of the committee that oversees the National Institutes of Health, said it would hold hearings on the commission's findings. "I am gravely disappointed that the commission did not recommend exploring substitutes to using human embryos for scientific experimentations," Bliley said in a statement.

India offers cut-price satellite launches

[LONDON] India promised last week to undercut the United States, Russia and China by 25 per cent in the market for commercial satellite launches, as it celebrated the launch of a rocket carrying three satellites. The Indian Space Research Organization says India can launch at least two satellites a year.

The Polar Satellite Launch Vehicle C2 blasted off from the island of Sriharikota in the Bay of Bengal, carrying satellites from India, South Korea and Germany. At 900 kilograms, the Indian satellite was the heaviest of the three, and will be used for oceanographic research.

Universities plan to fight threat of virtual learning

[LONDON] Leading universities in Britain, Australia, Canada and China are planning to form a 'super league' of institutions to meet the challenge to campus-based universities from distance-learning 'virtual' institutions. British vice-chancellors are particularly concerned about the threat posed by Internet-based programmes from leading US universities.

The super-league plan is understood to have been discussed at a meeting in Hong

Kong, hosted by the Association of University Presidents of China. One proposal is to give undergraduate students credit for courses taken at other universities in the league.

Israel and UK start up joint research fund

[JERUSALEM] A joint Israel-United Kingdom research and development fund will be established under an agreement signed last week by British industry secretary Stephen Byers and Israel's ambassador to the United Kingdom, Dror Zeigerman.

The fund, to be called Britech, will receive £15.5 million (US\$25 million) over five years from the two governments. It will function along the same lines as BIRD, the United States-Israel fund that provides grants for joint research and development projects.

Online journal launch could start price war

[WASHINGTON] A new chemistry journal was launched last week in a bid by the American Chemical Society and a consortium of academic libraries to force down the price of a far costlier competitor. *Organic Letters* will be published in both print and online versions.

The journal has support from the Association of Research Libraries, through

its Scholarly Publishing and Academic Resources Coalition (see *Nature* 398, 272; 1999). The new journal will cost \$2,300; its most expensive competitor, *Tetrahedron Letters*, published by Elsevier, costs \$8,602.

San Diego schools lose science and maths grant

[SAN DIEGO] The US National Science Foundation (NSF) has cancelled a \$15 million grant to the San Diego Unified School District. After a mid-point review of the five-year award made in 1996, the NSF notified the district last month that it had failed to meet the requirements of the Urban Systematic Initiative, which is designed to enhance science and mathematics teaching.

The withdrawal of the grant reflects the difficulty of improving education in urban areas. Three such NSF grants have previously been cancelled in New Orleans, Cleveland and Cincinnati, and a programme in Milwaukee is being placed on probation.

Buried nuclear waste 'must be retrievable'

[LONDON] Radioactive waste should be stored underground, but must be retrievable, according to a public panel participating in Britain's second national

consensus conference, which ended last week. The panel added that nuclear power should not be expanded, and no new contracts taken up to reprocess spent nuclear fuel until a solution is found to the problem of dealing with radioactive waste.

The report has been welcomed by the nuclear industry, environmental groups and the government. Environment minister Michael Meacher promised to pay careful attention to the recommendations. Ministers effectively ignored the recommendations of the first consensus conference five years ago, on plant biotechnology, in which a lay panel backed the technology, but called for stricter monitoring (see *Nature* 372, 122; 1994).

Russia's science minister keeps cabinet post

[MOSCOW] As widely anticipated, Russia's minister for science and technology, Mikhail Kirpichnikov, has kept his post in the cabinet of the new prime minister, Sergei Stepashin, whose members were announced last week (see *Nature* 399, 193; 1999).

Another survivor is Yevgeniy Adamov, who remains as minister responsible for atomic energy. But one departure is Vladimir Bulgak, who had been a vice-prime minister in a series of governments, and who included science among his responsibilities.



The full text of items on this page — and further details and background information about the Unesco/ICSU World Conference on Science, to be held in Budapest from 26 June to 1 July — can be found on *Nature's* website at <http://www.nature.com>

Internet gateway is planned to help women in science

[LONDON] An Internet-based gateway for women in science is to be officially launched at the World Conference on Science. The Gender, Science and Technology Gateway is a web-based clearing house of information, case studies and resources on gender and science and technology for sustainable development.

The gateway is a project of the Gender Advisory Board of the UN Commission on Science and Technology for Development in collaboration with the Once and Future Action Network (OFAN).

The advisory board is hoping to improve the flow of information to decision-makers on gender, science and technology. Secretariats in Indonesia and Montevideo are creating the gateway to disseminate relevant knowledge to developing countries, and to others wanting to gain access to the website.

The site will be hosted by Women in Global Science and Technology, a group working to improve global networking among women scientists and technologists on critical issues for development.

A 'virtual pavilion' — which will highlight activities around the world in gender, science and technology — is also being arranged for the conference by OFAN. The pavilion will include computer terminals with web access.

The project's organizers hope the pavilion will highlight women's contributions in science to researchers, agencies and donors.

See also: <http://helix.nature.com/wcs/a38.html>

Science partnerships 'must be of benefit to all sides'

[LONDON] Poor countries often learn little from international research partnerships that are linked to the priorities of the developed country, or are dominated by its researchers, according to a report from a Netherlands development think-tank.

In contrast, they can gain much from North-South research networks that offer clear benefits to all participants, and not just developing-country partners, says the report from the European Centre for Development Policy Management (ECDPM).

The report adds that the most successful partnerships are self-funded, either through membership fees or income from contract work, and do not rely unduly on support from aid agencies. It recommends that North-South research activities could learn from the experience of successful research partnerships between developing countries.

It also points out that research partnerships have been slow to use the Internet to share information, partly because Internet connections in some of the poorest countries are not well developed.

The report's authors are Louk Box, director of ECDPM, and Rutger Engelhard, an independent consultant. They wrote it for a meeting of the Geneva-based United Nations Commission on Science and Technology for Development last month.

More effective research partnerships is one of several priority areas on which the commission is working. Others include biotechnology for development and a 'vision

statement' on the future of science and technology for development.

The report recommends that research partnerships should be organized around a common scientific interest that provides researchers from developed countries with a stimulating intellectual challenge, but is also "strongly embedded in the Southern social, economic and cultural context".

For example, the European Tropical Forest Research Network, established by the European Commission in 1991, contributes to the conservation of forests. But it also enables scientists to exchange information on tropical forestry research.

"The identification of a concrete, widely shared problem or goal is one of the key pillars supporting networks," says the report. "Networks that fail to develop such a focus do not survive their infant years."

In addition, the report says that Southern partners need to be involved in the management of the research. In return, Southern members must give a "strong commitment" to making the partnership a success.

The report says much has improved since the 1960s and 1970s, when donor agencies from the richer countries engaged in scientific collaboration largely "to implement their own agendas". But it says that changes are difficult to bring about within existing partnerships as "Southern partners are often insufficiently organized to collectively assess their needs and present their agendas".

Full text: <http://helix.nature.com/wcs/a37.html>

Why the South needs the North to help set up research networks

[TRIESTE, ITALY] Scientists from Northern countries should actively help their colleagues in the South to work together in developing the networks and research infrastructure needed to increase their scientific capabilities, according to José I. Vargas, Brazil's minister of science and technology between 1992 and 1998.

Vargas, who is president of the Third World Academy of Sciences and the Third World Network of Scientific Organizations, says that such 'Southernization' of the North's scientific agenda is likely to be the most effective way to tackle problems such as climate change and food production, and to ensure that advanced science is brought to bear on the problems of poorer nations.

"Scientists in the North do not have to



look far beyond their laboratories to realize that global scientific knowledge is far more evenly distributed than the enormous wealth that scientists and technologists have created through their efforts," says Vargas. "Many of the North's graduate schools in basic sciences would find it difficult to maintain current levels of scholarship and research without a steady flow of graduate students from abroad."

He points out that 'South-South' cooperation has become one of the guiding principles of science and technology policy throughout the developing world. "Such efforts could be significantly strengthened

through programmes that facilitate North-South cooperation — provided that Northern scientists are responsive to the issues raised by colleagues in the South, and that scientists in the South share knowledge they have gained with their counterparts throughout the developing world."

He cites the way that Brazil was able to use scientific assistance from the United States to set up its space programme — and now not only uses satellite technology to tackle issues ranging from atmospheric carbon dioxide concentrations to soil and water quality, but also shares the information gathered from satellites with scientists in other developing countries.

Full text: <http://helix.nature.com/wcs/c19.html>

See also: <http://helix.nature.com/wcs/c18.html>

Cautionary tale on safety of GM crops

Sir — I am very much in favour of *Nature's* use of Scientific Correspondence to bring to our attention potentially important research before full publication of results. But there is a need for scientific rigour in the presentation of the information to ensure that it is not misrepresented. The report by Losey *et al.* in Scientific Correspondence is a preliminary research finding offering insight into the topical and important issue of genetically modified (GM) crops (*Nature* 399, 214; 1999). It has alerted the UK Advisory Committee on Releases to the Environment (of which I am chairman) to a potential problem that will require very serious thought.

It has also been assumed to demonstrate that GM crops harm butterflies and has fuelled public anxiety about such crops. The report is highly indicative of harm caused by pollen from the maize plants used

in this particular study, but the work is at this stage preliminary rather than definitive. It is clearly stated in the text, for example, that pollen for the non-*Bt* (*Bacillus thuringiensis*) maize control was from an unrelated, untransformed hybrid. However, there is no reported control to demonstrate that pollen from the transformed variety was not toxic in the absence of the functional *Bt* gene.

Losey *et al.* refer to a study (their ref. 3) stating that maize pollen is dispersed over at least 60 metres, and comment that a substantial portion of available milkweed (eaten by monarch butterflies) may be within the range of corn pollen distribution. But Losey *et al.* do not report a dilution study, so there is no evidence that dilute pollen would have an impact on larvae. Of course it is desirable to point out the potential harm that may arise from

pollen dispersal, which in this case could be very important, but the data reported by Losey *et al.* do not directly pertain to this issue.

On 20 May I stated on the BBC Radio *Today* programme that the letter by Losey *et al.* was not peer reviewed and that the work might be flawed. I am now aware that it was peer reviewed (as are all contributions to Scientific Correspondence), and wish to apologize. My suggestion that the work might be flawed was not intended as a slight but was a reminder to the press that preliminary observations should not be overinterpreted. Regrettably, most reporting of the communication has almost entirely ignored the need for such caution.

John E. Beringer

School of Biological Science, University of Bristol, Woodland Road, Bristol BS8 1UG, UK

Plant DNA patents in the hands of a few

Sir — Consolidation in the agrochemical and seed industry continues to shorten the list of owners of 'enabling' intellectual property for plant genetic modification and molecular biology. Six major groups of multinational companies are poised to develop and patent agronomically important plant genes identified in genome sequencing programmes for crop improvement (see page 396 in this issue and ref. 1). The ensuing cascade of patenting will have important implications for technology access and global food security. We report below on the ownership of patent applications with claims for DNA sequences from the world's major crops.

The finding that most patent applications have been filed by industry suggests that a strong advocate for the developing world will be needed to establish partnerships between industry and public laboratories to ensure that the benefits of genetic-modification technology will be widely available. This should be broadly achievable within the current patent regime, but we urge patent holders to be aware of their wider responsibilities when considering licensing terms.

We analysed patent applications with a filing date between 1980 and 1996 that make claims to DNA sequences (GENESEQ and World Patent Index databases, Derwent Ltd). The search, limited to 78 plant species of economic or scientific importance, revealed 601 patent applications containing

one or more DNA sequences from 60 species. About half have been granted.

About three-quarters of the applications were filed by 115 companies. Half of this total (48%) were filed by 14 multinationals. Recent corporate acquisitions by Monsanto have resulted in the company having the largest stake, with 69 applications, followed by Zeneca and Novartis.

Twenty-eight per cent of applications were filed by public-sector institutions. Over half of these are US owned. Four US universities and the US Department of Agriculture are in the top ten public institutions. In Europe, it is research institutes rather than universities that have sought to protect their inventions. Although the UK John Innes Institute and Germany's Max Planck and Institut für Genbiologische Forschung institutes have together filed half of the European public-sector total, Europe's stake is relatively low at a quarter of the world public-sector share. Public and private US organizations have 43% of the total, compared to 25% for Europe and 19% for Japan.

Maize was the most heavily patented species, claimed in almost 10% of the total, closely followed by the model plant *Arabidopsis*. About 40% of the patent applications included claims for cereals and pulses, one-third included fruit and vegetables, while claims to oilseed and model plants occurred in about one-quarter. Species used for fibres, beverages, herbs and spices accounted for the rest (10%). The overall total exceeds 100% because some applications include more than one species. Patent applications were not confined to crop species grown in developed countries.

Table 1 Predominant functions claimed for plant genes in the 601 patents analysed

Nutrition	109
Pathogen resistance	106
Regulatory DNA sequences	97
Growth and morphology	56
Sterility	45
Herbicide tolerance	36
Ex-plant production	30
Antisense	24
Plants as producers	22
Environmental	21
Colour	20

Some patents are counted in more than one category. A further 69 patents covered fibres, diagnostics, mapping and miscellaneous categories.

DNA sequences from nutmeg, cinnamon, rubber, jojoba and cocoa have been claimed by inventors from developed countries.

The applications, which rarely claimed more than one sequence, were dominated by genes associated with nutrition (20%), pathogen resistance (20%) and gene regulation (18%) (Table 1). Those in nutrition included genes that determine the amount or type of sugar, starch, oil or protein in the plant, encode enriched proteins in rice. Genes that contribute to pathogen resistance encode a wide variety of enzymes including chitinases. Regulatory DNA sequences in the form of transcriptional promoters are claimed that are in general tissue-specific. The gene sequences regulated by these promoters are also claimed in some cases.

Plant development and reproduction are of central importance to plant breeders, and 10% of applications fell within this class. Several concern modification of ethylene

production and the extent and pattern of flowering. Genes that confer male sterility are of particular value to the breeder, and account for about 8% of applications.

Although much of the controversy surrounding genetically modified crops concerns herbicide tolerance, only 7% of applications relate directly to this trait. These include genes encoding glutathione S-transferase IIIc, acetolactate synthase, lycopene cyclase and a protein conferring glyphosate resistance. The complexity of gene function is well illustrated by the acetyl-CoA carboxylase gene, which confers herbicide tolerance in monocotyledons but is claimed primarily for regulating oil content.

S. M. Thomas*†, M. Brady†, J. F. Burke‡

*Nuffield Council on Bioethics,

28 Bedford Square, London WC1 3EG, UK

†Science Policy Research Unit and

‡Department of Biochemistry,

University of Sussex, Brighton BN1 9QG, UK

1. Nuffield Council on Bioethics *Genetically Modified Crops: Ethical and Social Issues* (NCOB, London, 1999).

Papers should spell out authors' roles

Sir — When scientific papers with multiple authors are published, an indication should be given of the contribution that each author has made to the research. *Nature* recently published a paper by more than 20 authors. We can only guess what the contribution of author No.17 might have been, but it might be important to let the reader know about that, not least for the benefit of author No.17 himself. We propose the following system, which would take up minimal space in the journal.

All authors should be listed again at the end of the paper with a short statement about their contributions (Table 1). No matter if the author is the first or the seventeenth, they will be able to show their part in the work, which might be helpful for their future careers, especially in the case of younger scientists¹. This system will be informative for readers and also for potential employers, who need to assess the work of scientists they might employ.

This statement of authorship would also strengthen scientific teams, because it would become more difficult for someone to usurp rewards not belonging to him². Possible resentment about positioning

within the list of authors would be reduced, because each author would be given full credit according to their personal contribution. The scientific community would benefit — better research would emerge from teams that cooperated rather than behaving like packs of wolves. And other scientists would find it easier to contact the right person to ask about a specialized area of the research.

A. Zyzik, T. Goldmann

Research Center Borstel, Parkallee 1-40,

D-23845 Borstel, Germany

1. Tarnow, E. *Nature* **398**, 657 (1999).

2. Schiermeier, Q. *Nature* **397**, 640-641 (1999).

Scientists must bridge the communication gap

Sir — Lewis Wolpert is characteristically thought-provoking in his Commentary "Is science dangerous?"¹. But he misses the point.

Wolpert informs us that scientists have "specialized knowledge of how the world works that is not easily accessible to others". In fact, scientists show us an aspect of how the world can be considered to work. This is related to a particular way of thinking that has sufficient common principles to allow for a community of thought and discussion. The great power of this method is its predictive capacity and its potential for application in everything from space rockets to genetically modified crops.

But the scientific method does not, as Wolpert believes, "tell us how the world is". What gives scientists their special voice and power is not the 'truth' of their theories, but the application of these theories in technology.

Power is dangerous and therefore so, potentially, is science. Not because, as Wolpert scornfully suggests, our stupid culture is afraid of knowledge, but because scientists do not seem to be able to understand non-scientists. The understanding of scientists is limited to their particular approach to life. Or, to parody Wolpert, non-scientists have unspecialized knowledge of how the world works that is not available to scientists. Often the attempts of scientists to communicate to non-scientists only reinforce the divide: those who are interested in science enjoy the popularization; those who are not, do not.

Science, therefore, is dangerous because it is out of contact with much of its user base and, from some perspectives, is close to a tyranny. For too long scientists have patronized the non-scientific majority, and carried on with little concern for their reservations. The high-handed "It is essential to recognize" of Wolpert's article

belongs in the past. Scientists are boxing themselves into a corner by their inability to see that other people have a legitimate right not to see the world scientifically, and by their poor social skills. In science it is a virtue to be somewhere between forceful, condescending and arrogant. When communicating science, it is a disaster.

From 'mad cow disease' to genetically modified food, scientists have been failing to convince. No longer able to understand the language and aspirations of their fellow humans, they are moving from the position of curious outgroup to vulnerable minority. Science is useful, and the world it reveals is amazing. But Wolpert and the rest of us must understand that scientists can no longer dictate to the world. It is essential to realize.

Nick Battey

School of Plant Sciences,

The University of Reading, Reading RG6 6AS, UK

Sir — Wolpert's Commentary¹ seems quite naive. Take the key point of the supposedly 'neutral' knowledge provided by science. Wolpert says: "It is essential to recognize that reliable scientific knowledge has no moral or ethical value. Science tells us how the world is." From this, all the rest of the article's argument follows — we can't discuss reality, we can only accept it.

This crucial idea is in itself dangerous. It should be obvious that all knowledge has been acquired and is therefore a mix of 'reality' and our own way of understanding — the glasses with which we observe, and distort, reality. These 'glasses' include reductionism (see ref. 2 for simple examples in biology), and the necessity of building stable entities that can resist controversies ('black boxes'^{3,4}) and rapidly circulate within scientific networks³.

To use a crude analogy, science summarizes reality as much as a football score sums up two hours of emotions, missed opportunities and referee's mistakes. Any fan knows that the score does not exhaust the game, it only allows us to build a league table. Similarly, science chooses to extract from reality those features that allow it to build theories, and this demands high technology and a specific social organization.

The knowledge provided by science stems from the way the world is, but also from the way science has chosen to deal with it. Science is interwoven with technology, and the argument that 'science is pure, only its (technological) applications can be bad' might not be convincing for much longer in these distrustful times.

Pablo Jensen

DPM-Bat 203, 69622 Villeurbanne Cedex, France

1. Wolpert, L. *Nature* **398**, 281-282 (1999).

2. Lewontin, R. *Biology as Ideology* (Harper, New York, 1992).

3. Latour, B. *Science in Action* (Harvard Univ. Press, 1988).

4. Jensen, P. *Am. J. Phys.* (submitted).

Table 1 **Personal contributions to papers**

Author 1	Head of project, project design, coordination, initial idea
Author 2	General realization
Author 3	Reviewing
Author 4	Immunohistochemistry
Author 5	Electron microscopy

Deer destiny determined by density

Andrew Cockburn

Theory predicts that socially dominant females will produce more sons than daughters. But when times are hard, the higher growth rate of males means that proportionately more will die during development. New work shows that, in red deer, these effects can operate in concert.

When friends or relatives have a new baby, we inevitably ask about its weight and sex and expect the answers to be followed by the qualitative observation that mother and baby are "doing well". The relationship between offspring condition and sex, and maternal condition, lies at the heart of one of evolutionary biology's most famous predictions — the contention by Trivers and Willard¹ that, in polygynous mammalian societies (where some males can mate with many females), mothers in good condition should produce sons, whereas mothers in poor condition should produce daughters. This hypothesis has been difficult to test, but Loeske Kruuk and her colleagues² have now used superb long-term data from red deer (*Cervus elaphus*; Fig. 1) to cast the Trivers–Willard hypothesis in a new light. Their results, reported on page 459 of this issue, may help to explain the inconsistency among previous studies of mammalian sex ratios³.

The Trivers–Willard hypothesis is conceptually straightforward, as long as the mother produces one infant at a time. A mother can allocate a certain amount of resources to her child, and she should not breed if she does not have at least the minimum resources required by her baby. But what if she has more than this minimum? The optimum sex for her offspring will then depend on the relationship between the extra resources given to her child and the number of offspring the child produces, which, in turn, determines her number of grandchildren. The sex of the offspring is significant here because, in polygynous mammals such as antelopes, elephants, seals and some primates, sons will often have much more variable reproductive success than daughters.

Some sons gather large harems because they are more attractive or better fighters than other males. These sons have the potential to be exceptionally successful, as they can sire the offspring of many females each year. But, as a corollary, this may mean that other sons never acquire a mate. By contrast, most

daughters will be able to find a mate and rear young, but there is a limit to maximum productivity — their reproductive lifespan. So, according to the Trivers–Willard hypothesis, if the enhanced condition from heavy maternal investment is maintained by the infant into adulthood, such investment will be



Figure 1 Dense deer — the red deer (*Cervus elaphus*). Kruuk and colleagues² have found that the sex ratios of offspring are affected by population density.

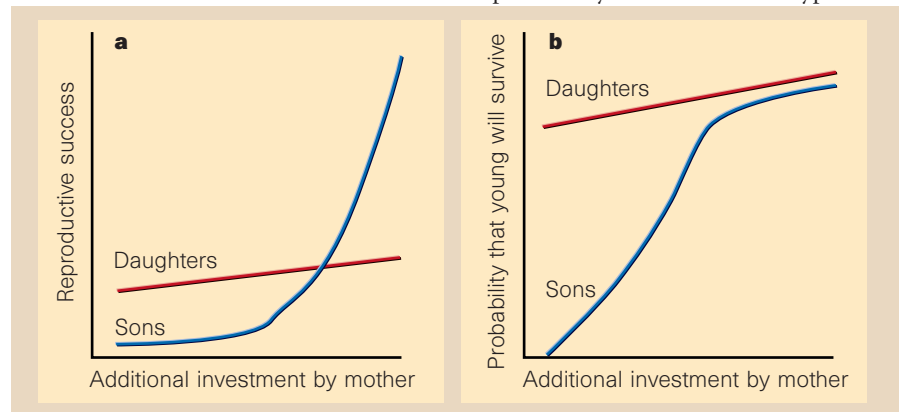


Figure 2 Two theories to explain the observed sex ratios in red deer populations on the island of Rum. a, The Trivers–Willard hypothesis states that, as mothers invest more heavily, they will gain a more rapidly increasing return (number of grandchildren) from sons than from daughters. So, there should be a point at which they switch from preferring daughters to preferring sons. In red deer, this point is associated with female dominance rank. b, An alternative hypothesis predicts that males may have a disadvantage early in their life because they grow rapidly, predisposing them to fetal mortality when resources available to their mother are in short supply. Attempts to breed by dominant red deer are unaffected by increasing density, which may predispose some high-ranking females to carry sons that die *in utero*.

most beneficial to sons. This is because a heavily provisioned son will have a good chance of enjoying the bonanza associated with holding a harem (Fig. 2a). Mothers in poorer condition, on the other hand, can allocate fewer resources to their offspring so should prefer the less variable return they obtain from daughters.

The most compelling support for the Trivers–Willard hypothesis comes from long-term studies of the red deer living on the Scottish island of Rum. Here, in the 1980s, Tim Clutton-Brock and his colleagues demonstrated all the separate elements of the hypothesis^{4,5}. Some female deer are dominant to others, and this dominance enhances the lifetime reproductive success of their offspring. However, higher rank increases the reproductive success of sons much more rapidly than that of daughters. As predicted, Clutton-Brock and colleagues found that dominant females are more likely to produce sons, whereas low-ranking females are more likely to produce daughters.

This influential work has inspired many studies of other species, from elephants and bison to mice and shrews. But support for the hypothesis has, at best, been inconsistent, or based on data that provide only some of the links in the chain of logic. Unfortunately, such partial support may be inconclusive. Even when almost all the links are present, we cannot necessarily assume that the chain will be completed⁶. Moreover, there are alternative explanations for certain sub-components of the Trivers–Willard hypothesis, which are not always acknowledged. For example, it is often reported that mothers in poor condition are more likely to have daughters, without showing that those in good condition produce sons. In this case, overproduction of daughters can be explained by an alternative hypothesis.

DAVID SHALE

Because large size assists males in the competition for females, rapid growth of males will be favoured by selection. So, sons may be more likely to die, either in the womb or during infancy, if their mothers find themselves in adverse conditions and cannot provide additional investment' (Fig. 2b).

Kruuk and colleagues² have now revisited the red deer on Rum. On certain areas of the island the population of red deer has increased because culling is not allowed. The authors found that, as the population density has increased in these areas, so the tendency of high-ranking females to produce sons has disappeared. Not only that, but the birth sex ratio as a whole is now biased towards females. A direct relationship to density seems likely because the correlation between condition of the mother and sex of her offspring persists elsewhere on the island, where culling is permitted and the population density remains low⁸.

How does this change with density relate to the Trivers–Willard hypothesis? It may be associated with increased vulnerability of male fetuses to mortality. In dense populations there will be increased competition for limited food and other resources, and, in red deer, the growth of male fetuses is much more sensitive to such changes in environmental conditions than that of female fetuses⁹. Kruuk and colleagues found that high-ranking females tend to reproduce each year, even in areas of high population density. So, more of these females may fall into the realm where fetal mortality is high (Fig. 2b), masking their tendency to conceive more sons. Such interactions may explain some of the inconsistent results in the literature. For example, Kruuk and colleagues point out that there has never

been positive support for the Trivers–Willard hypothesis from high-density populations of hoofed mammals.

Unsurprisingly, some questions remain. Why do dominant females breed every year, irrespective of population density? And what is the relationship between a mother's dominance and the lifetime reproductive success of her offspring now that density is high? Perhaps most important, the goalposts for tests of the Trivers–Willard hypothesis have been pushed back even further. The ability to match data on lifetime reproductive success with those on population density will be rare, except when studies persist through the lives of many generations of offspring — well beyond the term of a research grant or PhD. Indeed, where would the study of mammalian sex ratios stand if the red deer study had started when densities were already high, or it had terminated prematurely, before the effects of density could be appreciated? □

Andrew Cockburn is in the Evolutionary Ecology Group, Division of Botany and Zoology, Australian National University, Canberra, Australian Capital Territory 0200, Australia.

e-mail: andrew.cockburn@anu.edu.au

1. Trivers, R. L. & Willard, D. E. *Science* **179**, 90–92 (1973).
2. Kruuk, L. E. B., Clutton-Brock, T. H., Albon, S. D., Pemberton, J. M. & Guinness, F. E. *Nature* **399**, 459–461 (1999).
3. Clutton-Brock, T. H. & Iason, G. R. *Q. Rev. Biol.* **61**, 339–374 (1986).
4. Clutton-Brock, T. H., Albon, S. D. & Guinness, F. E. *Nature* **308**, 358–360 (1984).
5. Clutton-Brock, T. H., Albon, S. D. & Guinness, F. E. *Anim. Behav.* **34**, 460–471 (1986).
6. Krackow, S. *Behav. Ecol. Sociobiol.* **41**, 435–443 (1997).
7. Clutton-Brock, T. H., Albon, S. D. & Guinness, F. E. *Nature* **313**, 131–133 (1985).
8. Flint, A. P. F., Albon, S. D. & Jafar, S. I. *Gen. Comp. Endocrinol.* **106**, 374–383 (1997).
9. Post, E., Langvatn, R., Forchhammer, M. C. & Stenseth, N. C. *Proc. Natl Acad. Sci. USA* **96**, 4467–4471 (1999).

Display technology

Glowing developments

Karl Ziemelis

If you are a regular reader of *Nature*, the following description may sound familiar. A group of researchers have reported the discovery of an unusual physical phenomenon, the characterization of which promises new understanding of an important scientific question. At the end of the paper, however, the authors also raise the tantalizing possibility that the same phenomenon may find widespread application in a technological context. As so few of these promised applications seem to materialize, it could be argued that they represent little more than wishful thinking on the part of the authors. But occasionally a seemingly far-fetched claim of practical relevance does indeed live up to expectations,

and does so quickly. One of these success stories was celebrated, by both academia and industry, at a conference last month*.

The phenomenon under discussion was the electrically driven emission of light (electroluminescence) from non-crystalline organic materials — both polymeric and low-molecular-weight systems — the discovery of which was announced barely ten years ago^{1,2}. The proposed application is in flat-panel light-emitting displays, several impressive prototypes of which were exhibited at the meeting (Fig. 1).

Electroluminescence is usually harnessed in a device structure called a light-emitting diode (LED): essentially a thin film of a semiconductor (or several semiconductors) sandwiched between two electrodes. On applying a voltage, electrons enter from one electrode, positively charged 'holes'

from the other. Recombination of an electron and a hole, within the body of the semiconductor, produces a photon, and so leads to the emission of light. LEDs based on crystalline inorganic semiconductors have a long history, and their performance has improved steadily since their creation. It is when viewed in this context that the recent progress made with semiconducting organic materials seems so remarkable (Fig. 2).

Consider the electroluminescent polymers. The first² polymer LED emitted green-yellow light with a quantum efficiency (photons emitted per electron injected) of no more than 0.05%. Even when driven by a relatively high voltage (> 10 V), one needed to be in a darkened room to stand a chance of seeing the glow. Moreover, the early devices were extremely short-lived, usually lasting no more than a few minutes. Hardly a promising basis for the displays of the future! But only nine years later, operational voltages of < 5V, efficiencies of several per cent (the devices can be almost too bright to look at) and lifetimes of order 1,000–10,000 hours appear to be the norm. And the range of colours now available effectively spans the entire visible spectrum.

This remarkable advance in performance can be traced to progress on several fronts. First, to achieve the efficient — and balanced — injection of electrons and holes required to obtain a high photon yield, control of the electronic properties of the semiconductor–electrode interfaces has proved crucial (M. Schwörner, Bayreuth). This has invariably led to the introduction of at least one additional organic layer within the device structures, to facilitate the injection and transport of the charged species (usually the holes) against which the light-emitting material is predisposed. The introduction of such a layer has the additional advantage of moving the zone of emission into the body of the device, and so away from an electrode interface, where impurities can trap the charges and enable them to recombine non-radiatively. (See ref. 3 for a review.)

Second has been the marked increase in device lifetime, the key to which has been to understand and control the degradation processes within the organic materials. In the case of the polymers, systematic studies have helped to characterize the nature of any structural defects within the polymer chains resulting from undesirable reactions in the polymerization process; such information has then guided the development of improved syntheses (H. Becker, Covion). More generally, the expansion in material production from laboratory to industry scales has reduced chemical impurities to levels more commonly associated with the pharmaceuticals industry.

*Second International Conference on Electroluminescence of Molecular Materials and Related Phenomena (ICEL-2), Sheffield, UK, 12–15 May 1999.

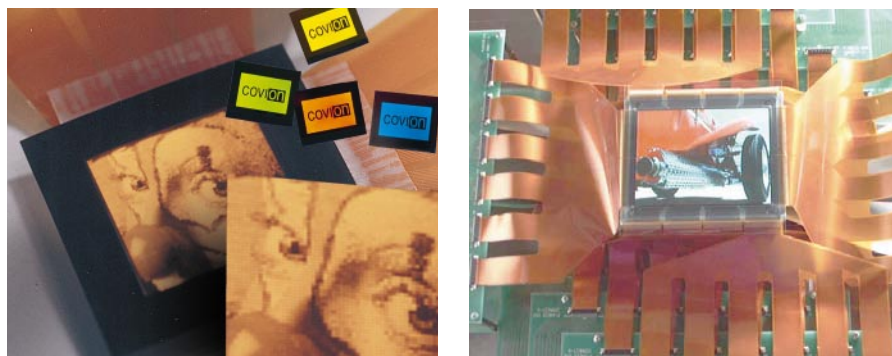


Figure 1 **Organics on display.** The left panel (main image) shows a prototype monochrome dot-matrix display fabricated from an electroluminescent polymer (Philips). Products based on this technology should be released next year. Also shown are four unpatterned polymer displays (essentially backlights), illustrating the broad range of colours that are available (Covion Organic Semiconductors). The right panel shows a prototype full-colour display based on small-molecule electroluminescence (Pioneer).

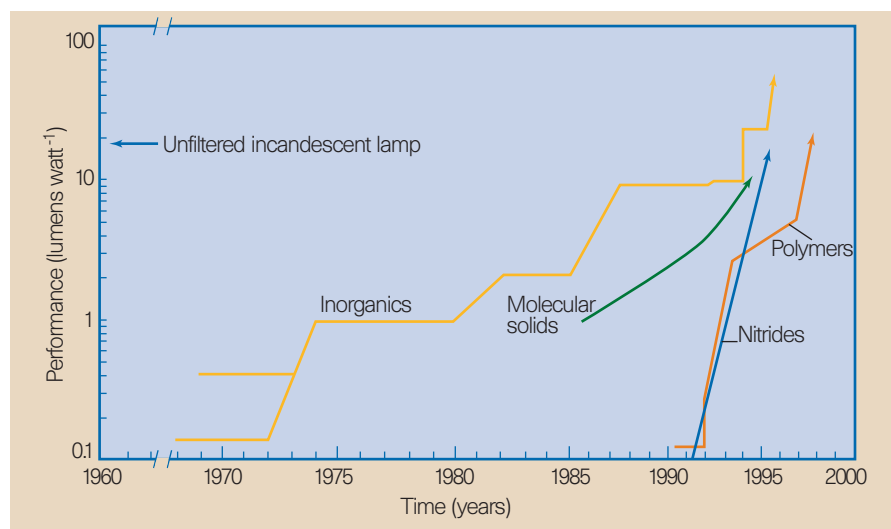


Figure 2 **Evolution of organic and inorganic light-emitting diodes (LEDs).** Performance is in lumens per watt (a typical value for an unfiltered incandescent lamp is shown for comparison). Inorganic LEDs based on III-V semiconductors, such as gallium arsenide, arrived on the scene in the early 1960s, and their performance has shown steady but significant progress since then (yellow curve). Organic LEDs (green curve for the small molecules; orange curve for the polymers) appeared much more recently, yet are already approaching the best performance levels of their inorganic counterparts. The nitride-based semiconductors (blue curve) are another success story², and have extended applications of inorganic LEDs into the blue-violet part of the spectrum.

(The introduction of zone-refining for the purification of inorganic crystalline semiconductors had a similar impact.) And chemists continue to design and synthesize new materials, with improved electrical and optical properties and overall stability: polyfluorene-based polymers look particularly promising (I. S. Millard, Cambridge Display Technology; M. Inbasekaran, Dow Chemical Company).

These improvements in efficiency and lifetime are already sufficient for the simpler applications (Fig. 1). For the polymers, several companies are well on the way to having products on the market, with monochrome backlights and segmented displays — such as those found on mobile telephones — promised by the end of this year (R.-J. Visser, Philips). High-resolution dot-

matrix displays based on the small-molecule systems should be available even sooner (M. Shimura, Pioneer).

But the Holy Grail of virtually all display technologies is full colour, and the organics are no exception. With an ever-growing palette of materials at their disposal, researchers are racing to find ways to combine them effectively into a single display. To do so poses formidable challenges. The materials are generally better suited for the deposition of large-area, uniform layers rather than intricately patterned arrays of differently coloured pixels; and the emission colours (and efficiencies) of the individual pixels need to be carefully balanced to match the spectral response of the eye. For the molecular systems, vacuum deposition through a patterned 'mask' provides a



100 YEARS AGO

From the hottest to the coldest stars I have found ten groups so distinct from each other chemically that they require to be dealt with separately as completely as do the Cambrian and the Silurian formations. ... I have gone further and defined the chemical nature of these stellar genera as the biologist defines the nature of any of his organic genera: we can say, for instance, that the Achernian stars contain chiefly hydrogen, nitrogen, oxygen and carbon, and to a certain less extent they contain proto-magnesium, proto-calcium, silicon and sodium, and possibly chlorine and lithium; so that at last, by means of this recent development of spectrum analysis, we have been able really to do for the various stars what the biologist a good many years ago did for the geological strata. ... My point is that the more one inquires into the chemistry of these things the more we come back to the stellar point of view and to the fact that, taking the simplicity of chemical form as determined by the appearance of these different chemical substances in the hottest stars ... and in relation to the "series" of spectra which they produce, we come to the conclusion that the first organic life was an interaction somehow or other between the undoubted earliest chemical forms.

From *Nature* 1 June 1899.

50 YEARS AGO

The question of drug resistance will require careful investigation Experiments at Entebbe indicate that strains of trypanosomes found in re-infected animals after the first treatment with 'Antrycide' may have considerable resistance to further treatment. This is illustrated by an experiment in which ten cattle were dosed with 2 gm. each of 'Antrycide' sulphate and were then exposed to trypanosome infection in a *Glossina pallidipes* area. Only one was alive six months later In some cases, animals were re-infected some months after treatment with 'Antrycide' and re-treated with curative doses of 'Antrycide' sulphate have developed cryptic but nevertheless fatal infections. Should such cryptic infections or prolonged latent infections prove to occur frequently after 'Antrycide' treatment, an accurate estimation of the true duration of the prophylactic effect will be difficult.

From *Nature* 4 June 1949.

promising route for achieving the necessary control, and a prototype five-inch display based on this approach was exhibited at the conference (T. Wakimoto, Pioneer). Solution-based polymers are not amenable to vapour-phase deposition, so alternative strategies have to be explored. The ink-jet technique (as used in colour printers) is one possible avenue for producing the necessary pixel arrays, although the initial results are far from optimal (H. Kobayashi, Seiko-Epson).

Of course, the display market is already well served by several more established technologies, so competition will be stiff. For the cruder monochrome applications, the combination of low power consumption, simplicity of fabrication and the 'gimmick' factor of having an unusually coloured display could well be sufficient to

give organic LEDs a competitive edge. The situation with full-colour displays is harder to predict and, in the shorter term, they will need to be shown to have a cost advantage if they are to succeed. But in the longer term, properties unique to the organic systems — not least their mechanical flexibility⁴ — may come to the fore, and it is conceivable that we may one day have colour displays that we can roll up and put in our pockets. Either way, electroluminescent organic displays in some shape or form should soon be seeing the light of day. □

Karl Ziemelis is physical sciences editor of Nature.
e-mail: k.ziemelis@nature.com

1. Tang, C. W. & Van Slyke, S. A. *Appl. Phys. Lett.* **51**, 913–915 (1987).
2. Burroughes, J. H. *et al.* *Nature* **347**, 539–541 (1990).
3. Friend, R. H. *et al.* *Nature* **397**, 121–128 (1999).
4. Gustafsson, G. *et al.* *Nature* **357**, 477–479 (1992).
5. Ponce, F. A. & Bour, D. P. *Nature* **386**, 351–359 (1997).

Apoptosis

Key to the mitochondrial gate

Jean-Claude Martinou

The mitochondrion, which was once thought simply to generate energy for a cell, is, in fact, a pivotal decision centre — it controls life and death by releasing death-promoting factors into the cytosol. One of these factors is cytochrome *c*, a protein that normally shuttles electrons between protein complexes in the inner mitochondrial membrane. But once released, cytochrome *c* helps to activate the caspases, a family of killer proteases. Release of cytochrome *c* from the mitochondria is controlled by proteins of the Bcl-2 family: those that prevent death (Bcl-2 and Bcl-x_L) inhibit release of cytochrome *c*, whereas those that promote death (Bax and Bak) induce this release. How does cytochrome *c* escape from the mitochondria? And how do members of the Bcl-2 family control this event? On page 483 of this issue, Shimizu *et al.*¹ show that Bax and Bak stimulate opening of the voltage-dependent anion channel (VDAC), a mitochondrial channel through which the cytochrome *c* permeates.

Two competing models have been proposed to explain how cytochrome *c* is released from mitochondria. In the first, a 'megachannel' called the permeability transition pore (PTP) opens^{2,3}, allowing water and solutes to enter. The mitochondrion swells, its outer membrane ruptures, and the mitochondrial proteins, including cytochrome *c*, escape. The PTP seems to consist of proteins from both the inner mitochondrial membrane (including the adenine nucleotide translocator, ANT) and from the outer membrane (VDAC). These proteins are thought to cooperate at sites where the two mitochondrial membranes are apposed, to form a large conductance channel that

allows molecules with a relative molecular mass (*M_r*) of less than 1,500 to pass.

This model accounts for some — but not all — types of cell death. For example, in many types of apoptosis the mitochondria remain morphologically normal. So other mechanisms might exist. Enter the second model, which predicts not the gross disruption

of the outer mitochondrial membrane, but the formation of a channel that is large enough to allow cytochrome *c* to pass through³. This model fits better with the absence of mitochondrial morphological changes. Given that cytochrome *c* is rather inflexible, owing to its haem prosthetic group, the diameter of the channel needs to be larger than that of cytochrome *c* (roughly 3 nm).

The ability of some Bcl-2-family members to form ion channels in artificial lipid membranes⁴, together with the fact that Bax and Bak trigger release of cytochrome *c* when added directly to mitochondria, has led to the hypothesis³ that Bax and Bak could be involved in forming such a channel. This hypothesis has now been tested by Shimizu *et al.*¹. By monitoring the movement of fluorescein-labelled cytochrome *c* through liposomes, they show that neither Bax nor Bak alone can form channels that allow passage of cytochrome *c*. But they provide compelling evidence that Bcl-x_L, Bax and Bak can interact with VDAC.

Also known as mitochondrial porin, VDAC is a small, abundant, integral-membrane protein. It is the main component (about 20%) of the outer mitochondrial membrane in eukaryotic cells^{5,6}. In planar lipid bilayers, when voltages of greater than 30 mV are applied, VDAC switches from high conductance (4 nanosiemens) and a large pore diameter (2.4–3 nm; 'open' state), to low conductance (2 nanosiemens) and a

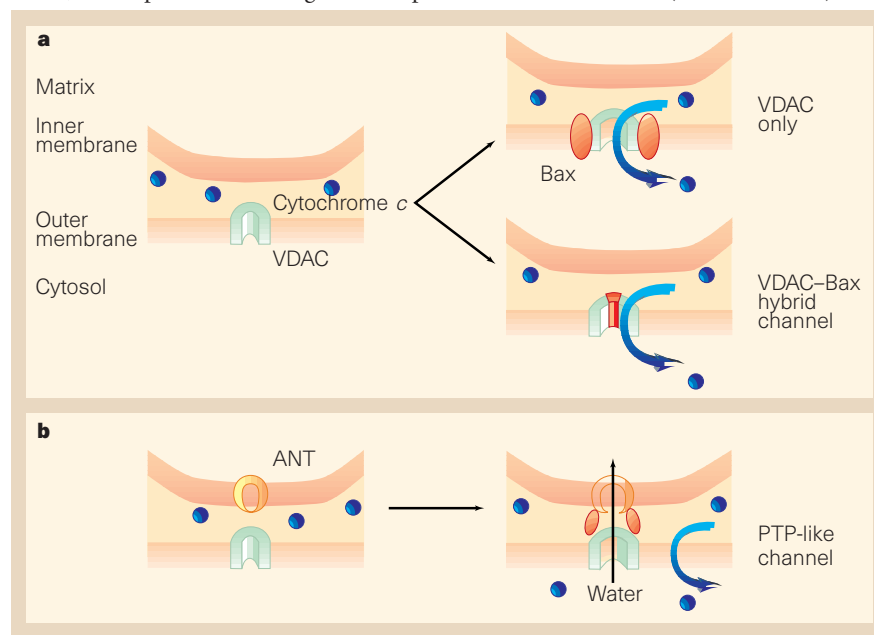


Figure 1 Models for the release of cytochrome *c* from mitochondria. a, The conformation of the voltage-dependent anion channel (VDAC) changes after interaction with Bax or Bak, which promote cell death. Possibly in combination with Bax or Bak, VDAC then forms a larger channel that is permeable to cytochrome *c*. This model is supported by the data of Shimizu *et al.*¹, and it is compatible with preservation of the outer mitochondrial membrane. b, Bax interacts with both the adenine nucleotide translocator (ANT) and VDAC to form a 'permeability transition pore (PTP)-like channel'. This channel allows water and solutes to enter the mitochondrion, causing the matrix to swell and its outer membrane to rupture. The result is release of cytochrome *c* and other proteins. Note that other mechanisms, which do not involve VDAC for cytochrome *c* release, are possible.

small pore diameter (2 nm; 'closed' state). Other agents that affect closure of VDAC include synthetic polyanions, cellular constituents such as NADH, and the so-called VDAC modulator⁶. Shimizu *et al.* now add Bcl-x_L, Bax and Bak to this list.

The authors reconstituted VDAC in liposomes, and show that Bcl-x_L stimulates closure of the channel, whereas Bax and Bak facilitate its opening. Moreover, they show that Bax and Bak allow cytochrome *c* to pass through VDAC. This is surprising, because the diameter of VDAC is normally too small to allow cytochrome *c* to pass. Shimizu *et al.* propose that, after VDAC interacts with Bax and Bak, its conformation changes. This allows VDAC — possibly in combination with Bax or Bak — to form a megachannel that is permeable to cytochrome *c* (Fig. 1a). The authors illustrated the requirement for VDAC using mitochondria purified from VDAC-deficient yeast mutants. When they added human Bax to these mitochondria, there was no release of cytochrome *c*. But by complementing the mutant mitochondria with human VDAC, they restored the ability of Bax to induce an efflux of cytochrome *c*.

In fact, VDAC is not the only protein required for the function of Bax in yeast — the ANT (ref. 7) and the F₀F₁-ATPase proton pump⁸ are also needed. Moreover, Bax can interact with the ANT, and Bax does not cause death in ANT-deficient yeast mutants⁷. Shimizu *et al.* claim that the ANT is not needed for release of cytochrome *c* through VDAC. But perhaps, under some circumstances, the ANT, through binding to Bax, may facilitate opening of VDAC. Such a picture would fit with the PTP opening model (Fig. 1b).

Another function of VDAC, in concert with the ANT, is ATP/ADP exchange — that is, it allows ATP to move out of the mitochondria and ADP to move in. In its closed conformation VDAC is impermeable to ATP⁶, and, earlier this year, Vander Heiden *et al.*⁹ reported that an early event in apoptosis (before cytochrome *c* release) is a defect in mitochondrial ATP/ADP exchange. So perhaps, during apoptosis, the ANT or VDAC (or both) fails to transport adenine nucleotides. In cells rescued by overexpression of Bcl-x_L, however, ADP/ATP exchange is stimulated to sustain coupled respiration⁹. In light of Shimizu and colleagues' results we can exclude the possibility that VDAC is responsible for this increase. Instead, it seems that Bcl-x_L closes VDAC, but that it maintains ADP/ATP exchange through a VDAC-independent mechanism.

Members of the Bcl-2 family are multifunctional — control of cytochrome *c* release is only part of their activity. They are found in other intracellular membranes, such as the endoplasmic reticulum and nuclear membranes, raising questions about whether they might control transport of

molecules across other membranes. Their mitochondrial activity may, however, be prevalent only in cells where the mitochondria are likely to be crucial to cell death. This may be the case in neurons, where, because of their mobility, mitochondria are ideal sensors of death signals that impinge on widely spaced regions such as the cell body, neurites and synapses. But wherever it may act, the arrival of VDAC into the apoptosis arena pinpoints this protein as a potential therapeutic target for preventing mitochondrial dysfunction in acute pathologies associated with apoptosis. □

Jean-Claude Martinou is at the Sero

Pharmaceutical Research Institute, 14 chemin des

Aulx, CH-1228 Plan-les-Ouates, Geneva, Switzerland.

e-mail: jean-claude.martinou@serono.com

1. Shimizu, S., Narita, M. & Tsujimoto, Y. *Nature* **399**, 483–487 (1999).
2. Kroemer, G., Zamzami, N. & Susin, S. A. *Immunol. Today* **18**, 44–51 (1997).
3. Green, D. R. & Reed, J. C. *Science* **281**, 1309–1312 (1998).
4. Kelekar, A. & Thompson, C. B. *Trends Cell Biol.* **8**, 324–330 (1998).
5. Mannella, C. A. *J. Struct. Biol.* **121**, 207–218 (1998).
6. Colombini, M., Blachly-Dyson, E. & Forte, M. in *Ion Channels* (ed. Narahashi, T.) 169–202 (Plenum, New York, 1996).
7. Marzo, I. *et al.* *Science* **281**, 2027–2031 (1998).
8. Matsuyama, S., Xu, Q., Velours, J. & Reed, J. C. *Mol. Cell* **1**, 327–336 (1998).
9. Vander Heiden, M. G., Chandel, N. S., Schumacker, P. T. & Thompson, C. B. *Mol. Cell* **3**, 159–167 (1999).

Climate change

Cornucopia of ice core results

Bernhard Stauffer

Natural archives of Earth's past climate take several forms — sea and lake sediments, tree rings, peat bogs and glacial ice — all of which are used in reconstructing climate history. But the records locked up in the large polar ice sheets are especially valuable. Cores of this ice not only allow reconstruction of changes in local temperature and precipitation, but also provide information about volcanic activity, storminess, solar activity and atmospheric composition.

Such records have already taken us back 150,000 years, a period covering about two glacial–interglacial cycles. Petit *et al.* (page 429 of this issue¹) now extend the most important records to four climatic cycles — that is, to about 420,000 years BP (before present). This extension has been made possible because, last year, ice-core drilling at Vostok station in Antarctica reached a record depth of 3,623 m. Most notably, analysis of the core allows investigation of whether transitions from glacial epochs to interglacials, and back again, always follow the same pattern or whether a variety of mechanisms is involved.

Over the past few years, ice cores from polar regions (Box 1) have delivered a variety of unexpected results. It is because of ice-core data that we know that large variations in climate were accompanied by naturally caused changes in the atmospheric concentrations of CO₂ and CH₄ — the most important greenhouse gases. Cores from Greenland provided the first evidence for fast and drastic climate changes during the last glacial epoch, including the transition to the present interglacial (the Holocene, the past 10,000 years) in the Northern Hemisphere. The records covering this transition, both from Greenland and from Antarctica, inspired various proposals as to the mechanisms causing or amplifying the temperature increase. With the extra data¹, these

proposals can now be tested further.

The four transitions from glacial to warm epochs, covered by the new Vostok records, started at about 335,000 years, 245,000 years, 135,000 years and 18,000 years BP. From this one would infer a roughly 100,000-year periodicity, and time-series analyses of the records indeed show a large, 100,000-year contribution to periodicity, along with another at 41,000-year intervals. This supports the idea that changes of the orbital parameters of the Earth (eccentricity, obliquity and precession of axis) cause variations in the intensity and distribution of solar radiation, which in turn trigger natural climate changes.

Of special interest is the interplay between greenhouse gases and climate. All four transitions from cold to warm climatic epochs have been accompanied by an increase in atmospheric CO₂ from about 180 to 280–300 p.p.m.v. (parts per million by volume; present concentration is 365 p.p.m.v.), and in atmospheric CH₄ from 320–350 p.p.b.v. to 650–770 p.p.b.v. (parts per billion by volume; present concentration 1,700 p.p.b.v.). Petit *et al.*¹ report that, within the uncertainties in the record, the increases in Antarctic temperature, CO₂ and CH₄ were in phase during all four transitions.

By contrast, based on measurements on the same core, Fischer *et al.*² have claimed that for the last three terminations there was a time lag of 500 to 1,000 years between the temperature increase and the CO₂ increase. The question of lags and leads in climate change is obviously a highly important one. But identifying a 500–1,000-year time lag is taking the current data and state of knowledge to its limits. Uncertainties stem not only from the limited sampling frequency but also from the problem of assigning dates to the air-containing bubbles in the core³ (air becomes enclosed in bubbles only at about

Box 1: Ice cores south and north

There are ice-drilling projects in both Antarctica and Greenland. Conditions are harsh, and stuck drills are an ever-present problem.

At Vostok, in Antarctica (pictured), the Soviets started deep drilling in 1980. A depth of 2,202 m was reached in 1985, when it became impossible to continue. A second hole had been started in 1984, and it reached a final depth of 2,546 m in 1990. In 1989, the project became a Russian–French–US endeavour, and the following year a third hole was started; it reached 2,500 m depth in 1992 (age of the ice at the bottom being about



200,000 years BP) and finally 3,623 m in 1998.

In central Greenland, the European GRIP core drilling reached bedrock in July 1992 at 3,028 m depth; likewise the US GISP-2 drilling in July 1993 (3,053 m depth). The undisturbed part of both cores covers the

past 105,000 years. A deep drilling at North-GRIP will extend the age scale to cover at least the last interglacial (135,000 years).

Other projects in Antarctica are being run by various national and international groups. In 1996 at Dome Fuji (East Antarctica) the Japanese reached 2,503 m; in January 1999, the US project on the West Antarctic ice sheet hit bedrock at 1,004 m. The European Project for Ice Core Drilling in Antarctica (EPICA) is at work at Dome Concordia (East Antarctica, present depth 786 m) and a second drilling will start shortly in Dronning Maud Land.

B.S.

ture is known (the sharp temperature increase preceding the start of intense ice melting in the Northern Hemisphere). A highly simplified course of events for all four transitions would then be as follows: first, changing orbital parameters initiated the end of the glacial epoch; second, an increase in greenhouse gases then amplified the weak orbital signal; third, in the second half of the transition, warming was further amplified by decreasing albedo caused by melting of the large ice sheets in the Northern Hemisphere.

Analyses of polar ice cores have added immensely to our knowledge of the mechanisms that govern global climate change. Not least, the results of these analyses provide tests for climate models intended to predict possible future responses to increasing concentrations of greenhouse gases. The results, however, are not easily come by — ice-core drillings in polar regions (Box 1) are long and difficult projects, with many pitfalls. The paper by Petit *et al.*¹ is further demonstration that the perseverance of all those involved in drilling projects — scientists, technicians and funding agencies — has ample rewards. □

Bernhard Stauffer is at the Physics Institute, Climate and Environmental Physics, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland. e-mail: stauffer@climate.unibe.ch

1. Petit, J. R. *et al.* *Nature* **399**, 429–436 (1999).
2. Fischer, H., Wahlen, M., Smith, J., Mastroianni, D. & Deck, B. *Science* **283**, 1712–1714 (1999).
3. Schwander, J. *et al.* *J. Geophys. Res.* **102**, 19483–19494 (1997).
4. Broecker, W. S. & Henderson, G. M. *Paleoceanography* **13**, 352–364 (1998).
5. Chappellaz, J. *et al.* *Nature* **366**, 443–445 (1993).

100 m below the snow surface, and so air and ice at the same level are of different ages).

Even if there does indeed turn out to be a time lag, CO₂ can still be an important amplifier for the temperature increase during the glacial–interglacial transition, which itself lasts several thousand years. However, whether amplification by greenhouse gases was responsible for 50% of the temperature increase, as Petit *et al.* speculate, also remains a hypothesis for the moment. Other amplification factors are relative humidity (water vapour is a greenhouse gas), surface albedo (changing ice cover and vegetation) and planetary albedo (changing cloud cover).

The causes and mechanisms of CO₂ increase at the beginning of the four transitions are also open to debate. From the Vostok results for the last two transitions, Broecker and Henderson⁴ concluded that the Southern Ocean is likely to be the main agent in regulating atmospheric CO₂. Similarities between CO₂ concentration and Antarctic temperature for the previous two transitions, as well as other parts of the record, add further support to the idea that the Southern Ocean does indeed have a key role. But although there are plenty of ideas about mechanisms linking events in the ocean to those in the atmosphere (changes in CO₂ solubility, phytoplankton productivity, iron fertilization and so on), there is no clear evidence to support any of them.

The Greenland ice cores revealed that fast and drastic temperature changes in the Northern Hemisphere are almost synchronous with fluctuations in CH₄ (ref. 5). Those

fluctuations are caused by variations in the extent and activity of sources (mainly wetlands in the tropics and in northern mean latitudes) which depend on temperature and precipitation rates. Petit *et al.* speculate that the CH₄ jumps in the first three transitions had the same cause as the most recent one, where the evolution of Greenland tempera-

Retroviruses

Closing the joint

John M. Coffin and Naomi Rosenberg

We are starting to understand much about how retroviruses integrate their DNA into host genomes. We know, for example, how the viral integrase carries out initial events in the process. But retrovirologists have tended to ignore the subsequent reactions, preferring to pass the buck onto 'cellular repair systems'. A report by Daniel *et al.*¹ in *Science* may now shed the first ray of light on these systems. They have found that a cellular damage-sensing system is implicated in completing retroviral integration.

Retroviruses integrate their DNA into that of the host as part of their replication cycle² — a feature that sets them apart from all other agents that infect multicellular organisms. A structure called the preintegration complex is formed from proteins of the incoming virus particle. Within this structure, viral DNA is produced from its RNA genome by the action of reverse tran-

scriptase, leaving a double-stranded DNA molecule with flush ends. Integrase, probably acting with a cellular DNA-binding protein³, then associates with the ends of the newly made DNA, and carries out two reactions at each end (Fig. 1, overleaf).

The first of these reactions is 3' cleavage. Two bases are removed from the 3' end of each strand, leaving a hydroxyl group. In the second (strand-transfer) reaction, the integrase catalyses a direct attack by that hydroxyl group on the target cellular DNA. These two reactions occur 4–6 bases apart, so, although each strand of the viral DNA is joined to its target, there is a 4–6-base gap as well as a two-base mismatch at each end. The reactions carried out by isolated preintegration complexes stop at this point, indicating that cellular repair systems are needed to finish the job. Clearly, if left unrepaired, the gaps would cause serious damage — the checkpoint systems of the

cell would presumably not allow DNA synthesis until such breaks were repaired.

Daniel *et al.*¹ have now addressed this issue by attempting to infect cells defective in a damage-sensing protein complex with several different retroviral vectors. This complex — DNA-dependent protein kinase (DNA-PK) — comprises a catalytic subunit (DNA-PK_{CS}) and two smaller proteins called Ku70 and Ku86. In one experiment, the authors infected cell lines derived from severe combined immunodeficiency (*scid*) mice, which are defective in DNA-PK_{CS}, with a vector derived from an avian retrovirus. The result was massive cell death by apoptosis within about 12 hours of infection, around the time expected for the first integration events.

The authors got similar results when they infected these cells with two unrelated retroviruses, but not when they used a virus deficient in integrase. Because the defect in *scid* mice affects the normal joining process by which immunoglobulin and T-cell-receptor genes are rearranged^{4,5}, Daniel and colleagues conclude that DNA-PK is needed to complete integration, probably by repairing the gapped DNA left by integrase. In support of this theory, when the authors infected cell lines deficient in Ku86 or XRCC4 — a protein thought to recruit a DNA ligase to sites of damaged DNA — they obtained similar results. And when they infected HeLa cells, they saw a modest induction of DNA-PK activity within 12 hours of infection.

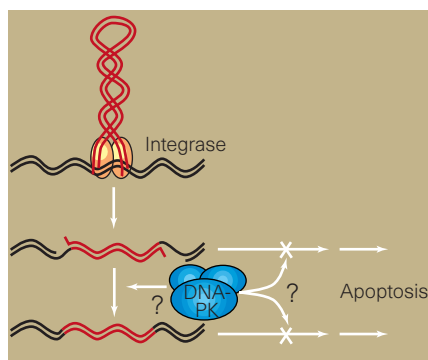


Figure 1 Possible involvement of the DNA-dependent protein kinase (DNA-PK) in retroviral integration. To insert their genetic material into a host's DNA, retroviruses use an enzyme called integrase. But integrase leaves a gap around the inserted sequence in the host's DNA — a gap that must be repaired before DNA synthesis can proceed. If the gap were left unrepaired, the cell would presumably die. Daniel *et al.*¹ now show that one of the host's damage-sensing systems involving DNA-PK is necessary to prevent this apoptosis. The DNA-PK could repair the gaps either directly or indirectly.

The idea that DNA-PK is directly involved in integration is an attractive but controversial one. Analyses of *scid* mice, and of mutant cells that cannot repair double-strand breaks induced by ionizing radiation, have established the involvement of DNA-PK in nonhomologous end-joining. But we do not know exactly how it interacts

with the broken ends and facilitates repair. The catalytic subunit is a large protein containing a carboxy-terminal catalytic domain that places it within the phosphatidylinositol 3-kinase family. Its activity is modulated by interaction with DNA and the Ku proteins, molecules that bind broken DNA. So perhaps DNA-PK shields the ends left at double-strand breaks, preventing them from being degraded. It could also facilitate repair by helping to bridge the gap between them, or by acting as a scaffold to recruit other components important for joining. But this system has not yet been shown to be directly involved in repairing single-stranded gaps such as those left by integrase.

Moreover, striking though they are, Daniel and colleagues' results have not been seen universally. For example, primary pre-B cells from *scid* or (DNA-PK_{CS} null) mice can be immortalized by infection with a retrovirus as readily as those from normal controls (ref. 6; N. Rosenberg and G. Taccioli, unpublished observation). Paradoxically, the pre-B cell lines used by Daniel *et al.* were derived in this way⁷. In addition, almost all *scid* pre-B cell transformants have managed to rearrange — albeit aberrantly — the immunoglobulin heavy-chain locus^{8,9}. At this stage, evidence for the role of DNA-PK in integration is indirect, and it will have to be established biochemically.

Given these considerations, perhaps DNA-PK protects the cell from a side-effect of integration, rather than acting directly in

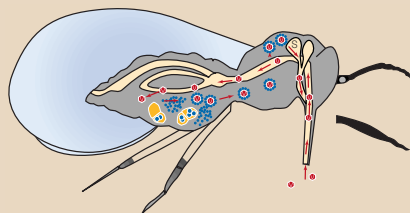
Parasitology

Chaperonin camouflage

Many viruses are transmitted between plants in the saliva of sap-sucking insects. But this is a risky business — billions of virus particles never make it. So, some viruses have evolved elaborate mechanisms to optimize their chances. The tomato yellow leaf curl (TYLC) virus, for example, is transmitted by whiteflies (*Bemisia tabaci*) in a way that involves passing through the insect's body. Reporting in *Virology* (256, 75–84; 1999), Shai Morin *et al.* show that, en route to the plant, the virus takes advantage of bacteria carried by the whitefly.

When a whitefly feeds from a plant infected with the TYLC virus (see diagram), virus particles (red) pass through the insect's gut wall and into the fluid (haemolymph) in its body cavity. The virus probably does not replicate in the insect, but, once inside, some viral particles reach the salivary glands (S). Here they are transported into the insect's saliva, from where they can infect another plant.

This strange mode of transmission has its advantages. Most other plant-infecting



viruses that use insect vectors are carried for only a few hours or days in the insect's mouthparts or throat. Their particles are shed when the insect moults its exoskeleton. Whiteflies that acquire the TYLC virus, however, often remain infectious for their entire life, increasing the odds of the virus being transmitted to another plant.

But passage through the insect is hazardous. Enzymes in the haemolymph degrade the virus particles, and this is where the bacteria help. *Buchnera* bacteria, which are found only in certain insects, live within specialized insect cells known as mycetocytes (yellow). These bacteria produce large quantities of a protein called GroEL (blue), some of which leaks into the

haemolymph. Although we don't know why GroEL is released, Morin *et al.* show that the virus particles avoid degradation by interacting with this protein. GroEL is a chaperonin — it binds to other proteins and helps them to fold. So it could change the conformation of the viral proteins, preventing them from being degraded.

Another kind of virus uses almost exactly the same survival trick. Potato leafroll virus is transmitted by aphids. As with TYLC virus, its particles pass through the insect's body, without replicating, from the gut to the salivary glands. Aphids also carry *Buchnera* bacteria, and the GroEL produced by these bacteria protects the viral particles. But whereas the potato leafroll virus has an RNA genome, TYLC virus has a DNA one, and their gene sequences and particles are very different. So these viruses have clearly converged on the same strategy independently. **Mark Gibbs**

Mark Gibbs is at the Research School of Biological Sciences, Australian National University, GPO box 475, Canberra 2601, Australia.
e-mail: mgibbs@rsbs.anu.edu.au

the reaction. Both the ends of the unintegrated DNA and the integration gaps themselves can, presumably, initiate an apoptotic signal through the p53 system⁴. Indeed, this signalling system may have evolved to protect the organism, not only from direct DNA damage but also from viral infection. Alternatively, some aspect of virus replication — such as expression of a viral gene — whose onset is temporally close to integration could be the initiating agent. In any case, this study should prompt retrovirologists to think more carefully about an important and understudied area of the virus–host interaction. □

John M. Coffin and Naomi Rosenberg are in the Departments of Molecular Biology, and

Microbiology and Pathology, respectively, Tufts University, 136 Harrison Avenue, Boston, Massachusetts 02111, USA.

e-mails: jcoffin_par@opal.tufts.edu
nrosenbe@opal.tufts.edu

1. Daniel, R., Katz, R. A. & Skalka, A. M. *Science* **284**, 644–647 (1999).
2. Brown, P. O. in *Retroviruses* (eds Coffin, J. M., Hughes, S. H. & Varmus, H. E.) 161–203 (Cold Spring Harbor Laboratory Press, 1997).
3. Lee, M. S. & Craigie, R. *Proc. Natl Acad. Sci. USA* **91**, 9823–9827 (1994).
4. Smith, G. C. M. & Jackson, S. P. *Genes Dev.* **13**, 916–934 (1999).
5. Bosma, M. J. & Carroll, A. M. *Annu. Rev. Immunol.* **9**, 323–350 (1991).
6. Fulop, G. M., Bosma, G. C., Bosma, M. J. & Phillips, R. A. *Cell. Immunol.* **113**, 192–201 (1988).
7. Schuler, W. *et al. Cell* **46**, 963–972 (1986).
8. Lieber, M. R. *et al. Cell* **55**, 7–16 (1988).
9. Malyon, B. A. *et al. Cell* **54**, 453–460 (1988).

Solar physics

Sunny side of global warming

E. N. Parker

In these times of speculation about climate change, the news that the Sun's magnetic field has doubled since 1900 — reported by Lockwood, Stamper and Wild¹ on page 437 of this issue — may not seem exciting. Neither are the implications immediately obvious. But it is a historical fact that our capricious climate responds to variations of the Sun's magnetic activity, with substantial warming and cooling with the rise and fall of activity over the centuries. The Sun is a complex and temperamental object, about which we know only the gross features. The magnetic fields produced by convection deep within the Sun are the basis for the complicated personality of our nearest and dearest star.

Lockwood, Stamper and Wild¹ use records of geomagnetic activity, monitored in England and Australia since the nineteenth century, to show that the weak (~ 0.5 millitesla) general magnetic field of the Sun has more than doubled over the past 100 years. The authors make use of the fact that the weak general field is stretched out through the Solar System by the expanding corona and solar wind. The impact of the stretched field against the magnetic field of the Earth is an important factor in creating geomagnetic activity. The new finding complements the well-known fact that the number of sunspots — cool, dark areas on the surface of the Sun with strong (0.2–0.3 tesla) magnetic fields — has also doubled over the same period of time (Fig. 1)². The general field varies much less with the 11-year sunspot cycle than does the number of sunspots, and appears to have a different origin. It is curious that the general level of both fields has increased by about the same degree over the past 100 years, suggesting some hidden commonality.

It is not yet clear what the doubling of the

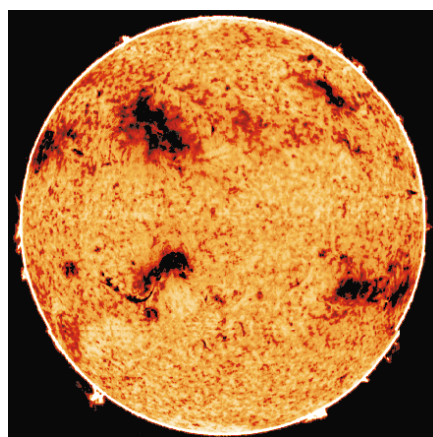


Figure 1 A spottier and more active Sun. The discovery by Lockwood *et al.*¹ that the general magnetic field of the Sun has doubled this century reinforces the likelihood that the Sun has had a hand in the general warming of our climate.

general field tells us about the detailed generation of solar magnetic fields, but together with the doubling of the sunspots it indicates a much more vigorous Sun, which, for reasons discussed below, is evidently brighter by around 0.1%. This seemingly innocuous brightening has occurred at the same time as a general warming of the Earth's climate and a substantial increase in atmospheric CO₂. The first thought is that the burning of fossil fuels at an accelerating rate since about 1950 is responsible for the increase in CO₂, whose greenhouse effect drives global warming. With fossil-fuel burning expected to rise, this line of thinking predicts a substantial desiccation of some geographical areas accompanied by melting of the polar ice caps and a rise in ocean level and flooding of vast coastal areas over the next century. The social, economic and political implications are fearful

to contemplate. So it is essential to check to what extent the facts support these conclusions before embarking on drastic, perilous and perhaps misguided plans for global action.

We begin with the fact that the total brightness, *B*, of the Sun has been monitored by radiometers on NASA spacecraft since 1978, which discovered that, astonishingly, the brightness of the Sun varies by an amount $\Delta B/B \approx 0.15\%$, in step with the 11-year magnetic cycle³. Monitoring other solar-type stars shows that this is normal behaviour, along with occasional shutdowns of activity for a century at a time (as during the Maunder Minimum of 1645–1715), or jumps to extraordinary levels of activity (as during the twelfth century Medieval Maximum). It is estimated that $\Delta B/B \approx 0.5\%$ during these abnormal centuries, the Sun being abnormally bright in the twelfth century and abnormally faint in the seventeenth. The mean annual temperature in the northern temperate zone of Earth was observed to vary by $\Delta T \approx 1\text{--}2^\circ\text{C}$ in association with these extended variations in solar activity and brightness⁴, and we cannot help noting that $\Delta T/T \approx \Delta B/B$.

In an ideal world one would turn to computer simulations of the terrestrial atmosphere to experiment directly with the climatic effects of varying solar brightness and increasing CO₂ in the atmosphere. But in the real world the atmosphere is too complex, with varying aerosol concentrations, varying wind patterns and ocean currents, deflection of wind by mountain ranges, strong vertical stratification of the atmosphere, and so on, to say nothing of the uncertainties in the formation of cloud cover. The numerical experiments carried out so far show that each of these factors plays an important role, but at present they cannot all be included simultaneously in a single numerical model, so there is no immediate answer to our questions.

Add to this the fact that most of the Earth's CO₂ is dissolved in the oceans and the rate at which the oceans remove CO₂ from the atmosphere depends in part on the surface seawater temperature. In the same way that a carbonated drink expels most of its CO₂ if warm and remains fizzy if cold, warmer seas mean less CO₂ is absorbed. Seawater temperatures have risen since 1900, with most of the warming of the atmosphere and oceans occurring before 1950 and most of the fossil-fuel burning after 1950. We know that sea-surface temperatures are influenced by the brightness of the Sun, so one wonders to what extent the solar brightening has contributed to the increase in atmospheric temperature and CO₂.

Finally, it has come to light that cloud cover follows the cosmic-ray intensity closer than any direct index of solar activity. We know that the condensation of water vapour

to form ice crystals and water drops is nucleated by ions in the atmosphere, and cosmic rays are the principal source of atmospheric ions. The intensity of cosmic rays is strongly suppressed by solar activity, so an increase in activity means reduced nucleation opportunities for cloud formation and an associated increase in sunlight reaching the surface of the Earth.

Ice cores from Greenland and Antarctica show that over the past 100 millennia there have been occasional precipitous drops in atmospheric CO₂, as well as large variations in mean temperatures. The mean temperatures are closely correlated with the general level of solar activity⁵. The inescapable conclusion is that we will have to know a lot more about the Sun and the terrestrial atmosphere before we can understand the nature of the contemporary changes in climate⁶. We expect that burning fossil fuel at the extravagant rate to which we have become accus-

tomed is a contributing factor, but so are the increased solar brightness and the increased seawater temperatures. In our present state of ignorance it is not possible to assess the importance of individual factors. The biggest mistake that we could make would be to think that we know the answers when we do not. Lockwood, Stamper and Wild¹ have contributed one more curious fact to our bin of knowledge, and we will need many more before all the mysteries are cleared up. □

E. N. Parker is in the Department of Physics, University of Chicago, Chicago, Illinois 60637, USA. e-mail: parker@odysseus.uchicago.edu

1. Lockwood, M., Stamper, R. & Wild, M. N. *Nature* **399**, 437–439 (1999).
2. Foukal, P. & Lean, J. *Science* **247**, 556–558 (1990).
3. Lee, R. B., Gibson, M. A., Wilson, R. S. & Thomas, S. J. *Geophys. Res.* **100**, 1667–1675 (1995).
4. Eddy, J. A. *Science* **192**, 1189–1202 (1976).
5. Stuiver, M., Grootes, P. M. & Braziunas, T. F. *Quat. Res.* **44**, 341–354 (1995).
6. Soon, W. H., Posmentier, E. S. & Baliunas, S. L. *Astrophys. J.* **472**, 891–902 (1996).

Quantum enzymology

Tunnel vision

Dagmar Ringe and Gregory A. Petsko

“The catalytic activity of thermophilic enzymes is low or absent at moderate temperatures at which conventional enzymes of similar function are optimally active.” Thus T. D. Brock¹, on what on the face of it is a surprising — but well characterized^{2,3} — fact. An understanding of why this should be so has been elusive. On page 496 of this issue⁴ Kohen *et al.* furnish evidence for an explanation that centres on a reduction in thermophilic (> 60 °C) enzyme flexibility at mesophilic (< 40 °C) temperatures.

Enzyme active sites are conserved, as are chemical mechanisms, so one would expect that homologous enzymes catalysing the same chemical reaction would do so at about the same rate at any given temperature. But they don't^{1–3}. One proposed explanation is that thermophilic enzymes might be cold-denaturing at lower temperatures; the crystallization and structure determination^{5,6} of many such enzymes at normal temperatures has, however, ruled that out. Another is that the mechanism changes in such a way that the rate-determining step is very different at different temperatures for enzymes isolated from thermophilic and mesophilic organisms; given the similarity in three-dimensional active-site structure between enzymes from both types of organism, this explanation also seems unlikely.

We and others have proposed that the explanation may instead lie in the temperature dependence of the dynamic properties of proteins. If atomic motions in the protein structure are essential for catalysis, and if many thermophilic proteins are more rigid

than their mesophilic counterparts at the same temperature, this difference in flexibility may account for the difference in reaction rate⁷. From an analysis of the temperature dependence of the reaction catalysed by a thermophilic alcohol dehydrogenase, Kohen *et al.* (a group in Judith Klinman's lab)⁴ now provide supporting data for this view. Most notably, the evidence comes not from the 'classical' rate component of the enzyme's hydride transfer mechanism, but from the component due to quantum-mechanical tunnelling.

In the classical model of chemical reactions involving particles surmounting a thermodynamic barrier, the reacting species must have enough thermal energy to reach the transition state. Later elaborations of this model include, for example, the viscosity-dependent diffusive motion over the top of the barrier, but the prediction that reaction rates should decrease continuously with decreasing temperature (Arrhenius behaviour) is common to all of them. However, the rates of some reactions become nearly independent of temperature at very low temperatures or show non-classical kinetic isotope effects^{8–10}. Quantum-mechanical tunnelling is a model that provides an explanation for these observations.

Tunnelling depends on the dual wave and particle nature of matter. If a barrier is sufficiently narrow compared with the wavelength of a particle, there is a finite probability that the particle can find itself on either side of the barrier without having to surmount it. (Tunnelling is like “winning the lottery without having to buy a ticket”¹¹.)

Because the wavelength of the particle does not depend on its kinetic energy, tunnelling is predicted to be temperature independent provided the width of the barrier does not depend on temperature. The de Broglie wavelength of matter is inversely proportional to the square root of the mass, so tunnelling has thus far only been inferred for reactions involving very light particles such as electrons and hydrogen species.

Enzyme-catalysed reactions have traditionally been treated as purely classical. This is not surprising, because tunnelling was thought to contribute substantially to barrier crossing only at very low temperatures, and enzymes are not normally studied at such temperatures. Work by Klinman and co-workers has suggested — controversially — that hydrogen tunnelling may in fact represent a significant component of the overall reaction in a number of enzymes even at room temperature^{12–14}. So certain enzyme reactions are being scrutinized for the possible involvement of tunnelling, including some that do not involve hydrogen as the tunnelling species.

Evidence for tunnelling in enzyme reactions has so far come from the magnitudes of primary and secondary kinetic isotope effects when deuterium or tritium is substituted for a hydrogen species that is transferred in the rate-determining step¹⁴. Temperature independence of a rate component has not been used as a criterion because the range of temperatures over which enzyme-catalysed reactions can easily be studied is small. Goldanskii has defined a ‘tunnelling temperature’ at which the contribution from tunnelling exceeds that from classical barrier crossing¹⁵. For hydrogen as the tunnelling species this is about 160 K (recall that room temperature is about 25 °C or 298 K). The hydrogen tunnelling temperature is lower than the ‘glass-transition’ temperature in protein dynamics, 200 K, below which many enzyme-catalysed reactions become so slow as to be almost quenched¹⁶. Below the glass-transition temperature collective motions in proteins are thought to have ‘frozen out’, and when this happens specific substrate binding at reasonable rates cannot be measured¹⁷.

The inability of enzymes to bind substrates and inhibitors below 200 K has been offered as evidence of the importance of protein flexibility for protein function. But although it is reasonable to expect that binding, which requires a molecule to fit into a cavity on the surface of a protein and displace any water therein¹⁸, would depend on the flexibility of protein structure, it has not yet been shown that protein motions are directly involved in the catalysis of bond making and breaking.

That is the issue that Kohen *et al.*⁴ set out to address. They chose to study a thermophilic enzyme because the severe tempera-

ture-dependent reduction in activity of the kind described by Brock¹ implies that there may be a dynamical transition in these enzymes at or near room temperature. The effect of this transition would be to freeze out certain protein motions that could facilitate hydride transfer (for instance, stretching motions along the reaction coordinate for C–H bond cleavage). Such motions might be important not only for classical barrier crossing but also for tunnelling, because the barrier width depends in part on the distance between the hydrogen donor and acceptor atoms¹⁹.

One issue raised by such an approach derives from the term 'mechanism', which implies a single, purely classical way of overcoming the transition-state barrier. Is tunnelling a different mechanism, or is it an alternative route within the same mechanism? We suggest that it is best considered as the latter — as one of two ways by which a thermal barrier can be crossed. Both are present in all reactions at all temperatures, but to different extents.

If protein mobility is important for both routes of barrier crossing, then the tunnelling component as well as the classical component of the rate should be temperature dependent. Moreover, in direct contrast to tunnelling behaviour in non-enzymatic reactions, where an apparently rigid barrier leads to increased tunnelling contribution at low temperature, the relative tunnelling contribution to the rate should decrease with temperature. That is exactly what Kohen *et al.* observe.

How does this result relate to the relative lack of activity of thermophilic enzymes at mesophilic temperatures? Kohen *et al.* find that the kinetic isotope effect is practically independent of temperature between 65 and 30 °C. But below 30 °C it displays a steep slope, implying that at around room temperature a dynamical transition occurs that impairs hydrogen tunnelling. No such transition is observed for the corresponding mesophilic protein. It is not yet known whether this transition corresponds to a shift in the glass transition from 200 K to 300 K, or whether, more likely, it is a second dynamical transition unique to thermophilic proteins. Determining this will require measuring the dynamic properties of thermophilic proteins over a wide temperature range, and not all thermophilic proteins may be appropriate²⁰.

Nevertheless, the tunnelling studies indicate that, at least in this particular thermophilic enzyme, rigidity at near room temperature impairs catalytic activity. Happily, the problem does not affect its function in the thermophilic microbe from which it was isolated — that organism never experiences such low temperatures in the wild. □

Dagmar Ringe and Gregory A. Petsko are at the Rosentiel Basic Medical Sciences Research Center,

Brandeis University, Waltham, Massachusetts 02454-9110, USA.

e-mails: ringe@brandeis.edu
petsko@brandeis.edu

1. Brock, T. D. *Science* **230**, 132–138 (1985).
2. Wrba, A., Schweiger, A., Schultes, V., Jaenicke, R. & Zavodsky, P. *Biochemistry* **29**, 7584–7592 (1990).
3. Schumann, J., Wrba, A., Jaenicke, R. & Stetter, K. O. *FEBS Lett.* **282**, 122–126 (1991).
4. Kohen, A., Cannio, R., Bartolucci, S. & Klinman, J. P. *Nature* **399**, 496–499 (1999).
5. Ursby, T., Adinolfi, B. S., Al-Karadaghi, S., De Vendittis, E. & Bocchini, V. J. *Mol. Biol.* **286**, 189–205 (1999).
6. Tanner, J. J., Hecht, R. M. & Krause, K. L. *Biochemistry* **35**, 2597–2609 (1996).
7. Zawodsky, P., Kardos, J., Svingor, A. & Petsko, G. A. *Proc. Natl Acad. Sci. USA* **95**, 7406–7411 (1998).
8. Bell, R. P. *The Tunneling Effect in Chemistry* (Chapman & Hall, London, 1980).

9. Saunders, W. H. *J. Am. Chem. Soc.* **107**, 164–169 (1985).
10. Northrop, D. B. *J. Am. Chem. Soc.* **121**, 3521–3524 (1999).
11. Goldanskii, V. I. *Sci. Am.* **254**, 46–52 (1986).
12. Cha, Y., Murray, C. J. & Klinman, J. P. *Science* **243**, 1325–1330 (1989).
13. Jonsson, T., Edmondson, D. E. & Klinman, J. P. *Biochemistry* **33**, 1471–1478 (1994).
14. Bahnsen, B. J. & Klinman, J. P. *Methods Enzymol.* **249**, 373–397 (1995).
15. Goldanskii, V. I. *Nature* **279**, 109–115 (1979).
16. Fitter, J. *Biophys. J.* **76**, 1034–1042 (1999).
17. Rasmussen, B. F., Stock, A. M., Ringe, D. & Petsko, G. A. *Nature* **357**, 423–424 (1992).
18. Ringe, D. *Curr. Opin. Struct. Biol.* **5**, 825–829 (1995).
19. Bahnsen, B. J., Colby, T. D., Chin, J. K., Goldstein, B. M. & Klinman, J. P. *Proc. Natl Acad. Sci. USA* **94**, 12797–12802 (1997).
20. More, N., Daniel, R. M. & Petach, H. H. *Biochem. J.* **305**, 17–20 (1995).

Evolutionary neurobiology

The neocortex comes together

Jon H. Kaas and Anton Reiner

A recent meeting* on the evolutionary developmental biology of the cerebral cortex brought together evolutionary biologists and comparative and developmental neuroscientists to discuss how the mammalian neocortex evolved. The neocortex is a thick, laminated sheet of neurons, which varies in size and regional differentiation between different mammals, and is especially impressive in humans. The question of how it evolved is challenging, because reptiles and other vertebrates have nothing quite like the neocortex, yet mammals and reptiles evolved from the same early amniotes (stem reptiles).

Living reptiles have a thin row of scattered neurons in their dorsal cortex, which has the same relative position as the mammalian neocortex but none of its size and complexity. Evolutionary neurobiologists would like to know how such a simple structure as the dorsal cortex was transformed into a large, cell-rich, laminated neocortex. Did other parts of the brain, specifically the dorsal ventricular ridge (DVR) of reptiles, also contribute to the evolution of the neocortex? The DVR is a cortex-like structure just below the lateral edge of the dorsal cortex (Fig. 1). It has some characteristics of the lateral (or temporal) part of the mammalian neocortex, such as inputs from the auditory and visual systems. So, the neocortex could have arisen from dorsal cortex alone, or from the dorsal cortex plus other parts of the forebrain such as the DVR.

An important premise of the meeting was that insights from developmental neurobiology help us to understand how the neocortex evolved. For example, changes in the timing and number of cell divisions in the ventricular proliferative zone could increase the sur-

face area of cortex. Then, outward migration of cells to the outer surface of the forebrain could increase its thickness¹. But this is just part of the process. Developmental studies clearly show that other parts of the brain also contribute to formation of the neocortex.

Neurons have been found, for example, to come in from a previously unsuspected source. Cells that generate a region below the cortex known as the basal ganglia, which is functionally associated with the motor system, also generate most or all of the inhibitory interneurons in the neocortex². These neurons reach the neocortex by migrating tangentially and dorsally from a lateral origin in the basal ganglia primordium. So, the neocortex is created by both outwardly and tangentially migrating neuron precursors (neuroblasts). As yet, however, there is no evidence that the cells that generate the DVR in reptiles contribute neurons by tangential migration to the neocortex in mammals. Moreover, evidence presented at the meeting, that early expression of a homeobox gene called *emx-1* differs between parts of the DVR and the temporal neocortex, argues that the DVR and lateral neocortex are not homologous structures.

As well as the migration of inhibitory interneurons to the neocortex, there is a second, tangential migration of cells from an adjacent part of the basal ganglia primordium. These cells provide some of the Cajal–Retzius cells in the surface layer of neocortex, as well as some of the sub-plate neurons just below it³. Cajal–Retzius cells, which also arise by local outward migration, induce the orderly formation of layers in the neocortex by releasing a molecule called reelin. So, an increase in the number of Cajal–Retzius cells might have been important for transformation of the dorsal cortex to neocortex. Sub-plate neurons seem to be involved in the formation of orderly thalam-

*Novartis Foundation Symposium on the Evolutionary Developmental Biology of the Cerebral Cortex, Wellcome Trust Auditorium, London, UK, 20–22 April 1999.

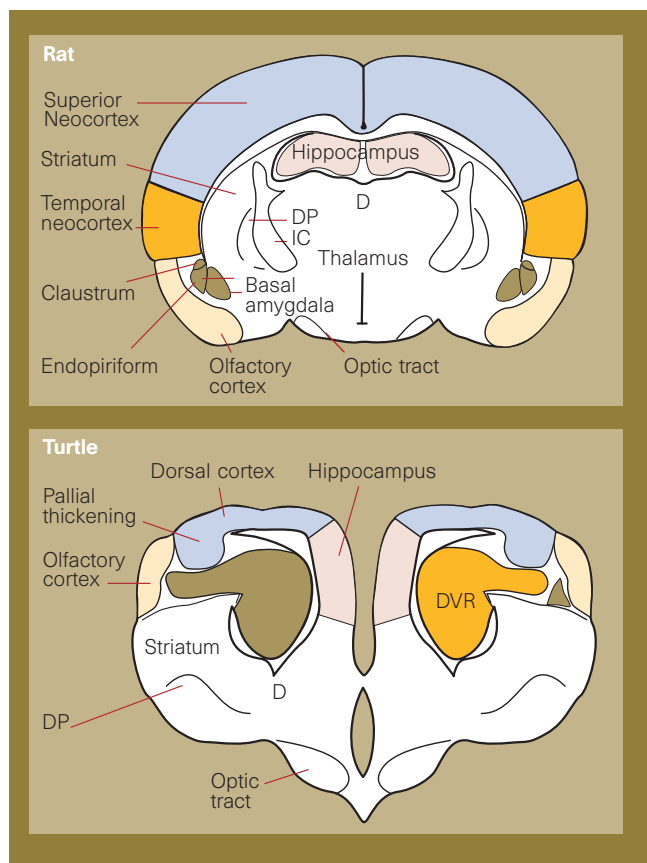


Figure 1 Two proposed courses for evolution of the mammalian neocortex and the forebrain of reptiles, from forebrain structures of a hypothesized stem amniote (from which both evolved). The main question is whether the temporal neocortex and the dorsal ventricular ridge (DVR) evolved from the same primordial structure in early amniotes (as depicted on the right of each panel) or not (depicted on the left). D, diencephalon; DP, dorsal pallidum; IC, internal capsule. (Modified from ref. 7.)

ic connections of the neocortex.

Modern developmental neurobiology is also helping us to work out how the neocortex is subdivided into its various areas. The neocortex varies greatly, not only in size — from a surface area of 1.5 cm² in small-brained shrews to 800 cm² per hemisphere in humans — but also in the number of subdivisions. Small-brained mammals typically have about 15 well-differentiated neocortical areas, whereas monkeys and humans have well over 50. We do not know how such subdivisions are formed, nor how their numbers increased during evolution. One possibility is that cortical areas develop owing to differences in gene expression, which, in turn, depend on position in the brain. For example, the main subdivisions of the forebrain are characterized by differences in the expression of homeobox genes. Such differences in expression can, in fact, be used to identify homologous portions of the forebrain across vertebrates⁴. So, regional differences in the expression of genes may differentiate the neocortex into areas, and gene duplication followed by gene divergence may increase the number of these areas during evolution.

Of course, neurons within the different areas need to form proper connections with other structures, and these connection patterns need to be modified as new cortical areas evolve. Perhaps growing outputs from the cortex interact with growing inputs from the thalamus as they approach each other, and appropriate matches are recognized

chemically. In this way, subsequent growth could be guided to the proper targets⁵ — an idea known as the 'handshake hypothesis'. Connection patterns could be altered by modifying these chemical-recognition signals. Another characteristic of developing nervous systems is that they initially generate patterns of neural activity that then further shape structural development of the nervous system⁶. To the extent that this occurs in the neocortex, this region could develop in structurally different ways depending on the factors that influence neural activity. And genetically mediated changes in the sensitivity of neurons, and the time course of that sensitivity to activity-dependent mechanisms in development, could be among the potent forces that shaped evolution of the neocortex. □

Jon H. Kaas is in the Department of Psychology, 301 Wilson Hall, Vanderbilt University, Nashville, Tennessee 37240, USA.

e-mail: Jon.H.Kaas@vanderbilt.edu

Anton Reiner is in the Department of Anatomy and Neurobiology, 855 Monroe Avenue, University of Tennessee, Memphis, Tennessee 38163, USA.

e-mail: areiner@utmem.edu

1. Rakic, P. *Trends Neurosci.* **18**, 383–388 (1995).
2. Anderson, S. A., Eisenstat, D. D., Shi, L. & Rubenstein, J. L. R. *Science* **278**, 474–476 (1997).
3. Lavdas, A. A., Grigoriou, M., Pachnis, V. & Parnavelas, J. G. *J. Neurosci.* (in the press).
4. Fernandez, A. S., Pieau, C., Repérant, J., Boncinelli, E. & Wassef, M. *Development* **125**, 2099–2111 (1998).
5. Molnár, Z. & Blakemore, C. *Trends Neurosci.* **18**, 373–379 (1995).
6. Shatz, C. J. *Proc. Natl Acad. Sci. USA* **93**, 602–608 (1996).
7. Reiner, A. *Comp. Biochem. Physiol.* **104**, 735–748 (1993).

Daedalus

Play it again, Sam

Last week Daedalus proposed taping a sequence of electrodes on the skin above a nerve trunk (the spinal cord, say). A voltage spike stepped upwards along the array at a specific velocity would, on travelling-wave-tube principles, fire the one fibre in the trunk with that exact transmission velocity. The array could thus address individual fibres, and induce controllable 'virtual sensations'.

He now muses that a voltage spike stepped down the spinal cord at a specific velocity would induce a pulse in the matching downgoing motor nerve. It would drive some muscle somewhere in the body. Detailed experiments would be needed to discover what nerve-fibres with what transmission velocities served which muscles; but then a controlled pattern of downgoing voltage spikes could drive just the right muscles in the correct sequence to induce a specific 'synthetic action'.

The obvious beneficiaries are dedicated joggers and exercisers. Once equipped with electrode arrays taped over the relevant nerves and driven from a portable computer pack, they would need no will-power. They could simply relax and let themselves be jogged along automatically by their own muscles. Assembly-line workers and others engaged on dreary repetitive tasks would also welcome nerve automation. They would find themselves repeatedly 'going through the motions' without thought or effort. Their automatic routines, optimized and pre-recorded, would be repeated with wonderful fidelity. Users could, of course, still exercise judgement, and vary their actions by added conscious intention.

Synthetic action would also be a splendid form of practical training. Skills such as typing or playing the piano could be programmed into an electrode array driving the subject's muscles, and replayed repeatedly. He would learn the 'feel' of the skill, and would soon be able to carry it out by himself.

The only snag is that a downgoing voltage spike, intended to activate a specific muscle, might also induce a downgoing impulse in some sensory nerve whose traffic goes the other way. This would cancel its upcoming signals and induce a selective anaesthesia. Nerve-automated personnel may be plagued with strange repetitive patterns of numbness. On the other hand, surgeons would welcome a tightly targeted local anaesthesia that could be instantly switched on and off.

David Jones

Obituary

Arthur Schawlow (1921–99)

Laser scientist

Arthur Schawlow, co-inventor of the laser and one of the great figures of this century in optical science, died on the 28 April 1999. The laser has revolutionized optical science and its applications, providing powerful new technology in a variety of areas such as communications, cutting and welding, information storage and precision measurements. Schawlow also carried out many innovative experiments of high precision and high resolution, and pioneered a variety of new experimental optical techniques.

Schawlow was born in Mount Vernon, New York, in 1921. His father had emigrated to the United States from Latvia, and his mother came from Canada. At her urging, the family moved to Toronto when Arthur was three. As an outstanding student, the young Schawlow was able to obtain a scholarship to the University of Toronto and proceed to a PhD degree. His thesis, supervised by Professor Malcolm Crawford, involved high-resolution spectroscopy using atomic beams, and the equipment required much ad hoc and money-saving improvisation. His later work was carried out in the United States — at Columbia University, Bell Telephone Laboratories and Stanford University, where he worked until shortly before his death aged 77.

In college, Schawlow also played the clarinet and helped to organize the Delta Jazz Band. Over his lifetime, he assembled a remarkable collection of jazz recordings, now in the Stanford archives.

After finishing his PhD, Schawlow went to Columbia University in 1949 to work with me on microwave spectroscopy, supported by a fellowship. I was quickly impressed by Arthur's insight and creativeness. He worked on a variety of

spectroscopic projects, including the first measurement of the microwave spectrum of a free radical, OH, a measurement which allowed OH to become the first molecule detected in interstellar space by microwave techniques. He even applied X-ray fine-structure results to a determination of charge distribution within atomic nuclei. During the early part of his career, at the University of Toronto and Columbia University, Schawlow contributed substantially to determinations of nuclear moments and properties. Arthur also accepted my invitation to jointly write a book, *Microwave Spectroscopy* (published in 1955 by McGraw-Hill).

In 1951, Schawlow married my youngest sister, Aurelia Townes, and accepted a position at Bell Telephone Laboratories. There, he was busy with superconductivity and spectroscopy.

In 1954, colleagues and I at Columbia built the first maser, a device that produces and amplifies microwave radiation, hence the name 'microwave amplification by stimulated emission of radiation'. After recognizing late in 1957 that maser-type techniques could be extended to wavelengths as short as visible light, I talked with Arthur about this. He had been thinking along similar lines, so we joined forces and it was he who suggested using a simple cavity of two parallel mirrors. This beautiful idea allowed the isolation of single modes. Discussions and correspondence over the next six months resulted in our paper "Optical and Infrared Masers", published in late 1958, which initiated the explosion of work on lasers, with the first working laser being built in 1960. The word laser was adapted from maser by Columbia students and is now universally used. The basic laser patent resulting from this work was the property of Bell Telephone Laboratories.

In 1961, Schawlow accepted a position at Stanford

University as a professor of physics. There he developed many new spectroscopic techniques using lasers, particularly oriented towards high spectral resolution and high precision. They included saturation

spectroscopy, multiple-photon phenomena and optical measurement of the Lamb shift. The many excellent students and postdoctoral physicists who worked with him included Theodor

Hänsch, who with Schawlow first recognized and discussed molecular or atomic cooling with laser light. This initial idea has allowed the development of a remarkable new field of low-temperature physics, and has been applied by other physicists to produce new low-temperature states of matter.

In Schawlow's oral history, now available in research libraries, he singles out three of his scientific papers as particularly important: one on determination of nuclear size from hyperfine spectra resulting from work with Crawford at Toronto; the optical maser or laser paper; and the one on slowing down of atoms or molecules with lasers.

Schawlow's approach to science involved deep and intuitive understanding. He liked simplicity and found ways of doing powerful experiments with relatively simple elements, and of using perceptive but simple analyses of phenomena. He was renowned among friends and students for his charm and sense of humour on all occasions, and for amusing scientific demonstrations and tricks. The latter include making an edible laser of Jell-O, and using a laser to explode one balloon inside another to illustrate its selectivity.

Schawlow was also known for his research on communicating with autistic people, and for his contributions to helping them. His son is at an institution in California for people with autistic problems. The main building was donated by Arthur and named after his wife Aurelia Schawlow, who died in an automobile accident in 1991. Shortly before his own death, the institution was renamed the Arthur Schawlow Center because of his crucial role in its success.

In addition to the Nobel prize, awarded in 1981 for his "contributions to the development of laser spectroscopy", Schawlow received many honours. He also served as president of the American Physical Society, president of the Optical Society of America and on many boards and committees that serve science and society.

The remarkable variety of contributions Schawlow made to science is already the basis of many important developments and we can expect more to come. His character, friendliness and humour will also be long remembered. He is survived by his son Arthur K. Schawlow, his daughters Helen Johnson and Edith Dwan, and by five grandchildren.

Charles H. Townes

Charles H. Townes is in the Department of Physics, University of California, Berkeley, California 94720, USA.
e-mail: cht@ssl.berkeley.edu



Orchid pollination by sexual swindle

The flowers of *Ophrys* orchids mimic receptive females of usually only one pollinator species. Males of this species are attracted primarily by the odour of a flower and transfer the pollinia during so-called 'pseudocopulations' with the flowers^{1–3}. We have found that flowers of *O. sphegodes* produce the same compounds and in similar relative proportions as are found in the sex pheromone of its pollinator species, the solitary bee *Andrena nigroaenea*. Common straight-chain saturated and unsaturated hydrocarbons are the key components in this chemical mimicry, which seems to be an economical means of pollinator attraction.

Naturalists have long wondered why the shape and colour of the flowers of *Ophrys* are similar to those of certain insects⁴. The odour of a flower was found to be the most important signal triggering pseudocopulation behaviour². Although many odour components were produced by the flowers as well as by their pollinators^{5–7}, there has been no convincing identification of the compounds that trigger pseudocopulation. We have used gas chromatography–electroantennographic detection to identify physiologically active compounds, which were subsequently tested in the field for behavioural activity. The activity of synthetic mixtures was compared with that of the sex pheromone of female bees and the pollinator-attracting scent of *Ophrys* flowers.

We tested various odour samples of female bees and orchid flowers for their attractiveness to male bees (Fig. 1). Of all bee-odour samples tested, cuticle extracts elicited the most intense behavioural reactions from males, indicating that this is where the sex pheromone of *A. nigroaenea* is located. Extracts from the *Ophrys* flower labellum also elicited frequent copulation attempts by males (Fig. 1).

In the experiments using gas chromatography–electroantennographic detection, 15 components from the attractive odour samples of female bees triggered electroantennographic responses in male antennae (Table 1). The *Ophrys* flower extracts contained all except one of these compounds. The compounds were identified by gas chromatographic–mass spectrometric analyses as straight-chain saturated and unsaturated hydrocarbons with 21–29 carbon atoms. Both sample types contained these compounds in very similar relative proportions (Table 1). To investigate the behavioural activity of electrophysiologically active compounds, we performed behavioural tests with male bees. A blend of 14 synthetic hydrocarbons occurring in both the bee scent and the flower scent, mixed in accordance with the relative proportions

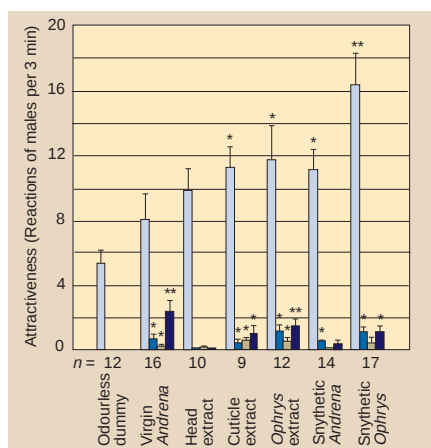


Figure 1 Reactions of *Andrena nigroaenea* males in behavioural tests in the field. Odourless dummies (Soxhlet-extracted, dried *A. nigroaenea* females) scented with different samples were offered to *A. nigroaenea* males in a patrolling area for 3 min. Means ($\pm$ s.e.m.) of approaches to the dummy to less than 5 cm (pale blue bars), pouncing on the dummy (royal blue bars), alighting on the dummy (beige bars) and copulation attempts with the dummy (dark blue bars) are plotted. The Mann–Whitney *U*-test was used for statistical analysis. Asterisks indicate significant differences between the reaction types of the odourless-dummy group and each of the other six test groups at significance levels of * $P < 0.05$ and ** $P < 0.001$.

found in either sample type, elicited frequent copulation attempts in males when applied to a dummy (Fig. 1).

Our understanding of the mechanisms of chemical mimicry in *Ophrys* orchids has been constrained by a limited knowledge of the chemical composition of sex pheromones in the solitary bees and wasps that serve as pollinators. Earlier studies reported aliphatic 1- and 2-alcohols and terpenes in flowers and bee odours to be the key components for pollinator attraction in several *Ophrys* species^{5,7}. However, these compounds did not elicit definite copulation

attempts in male bees. This led to the hypothesis that *Ophrys* flowers produce only a set of “second-class attractivity compounds” that attract only the part of the pollinator population with a low threshold for sexual stimuli⁵.

Our results challenge this view. In *O. sphegodes*, we find that the behaviour-inducing odour bouquet is more or less identical to the sex pheromone of the female bees and therefore elicits nearly equally intense reactions. But the behaviourally active compounds are ubiquitous or occur in at least trace amounts in plant cuticular waxes⁸. The similarities in their relative amounts in orchid labella and attractive female bees supports the view that the specific pattern of otherwise common compounds is the key innovation of *O. sphegodes* for the attraction and deception of the pollinators.

These findings shed new light on the evolution of this fascinating insect–plant relationship. Cuticular hydrocarbons, as a part of the plant surface wax, primarily prevent the loss of water. During the evolution of chemical mimicry in *O. sphegodes*, these compounds obtained the additional function of attracting pollinators, which was achieved through alterations in their relative proportions. It seems feasible that, in an ancestor of *Ophrys*, the pattern in a mutant occasionally resembled that of the sex pheromone of a pollinator species. This could have led to pollination by sexually excited males of this species, and natural selection would have favoured further plant mutants with a hydrocarbon pattern with an even closer resemblance. Apart from advantages discussed elsewhere⁹, being pollinated by pseudocopulating males allowed the plant to decrease the costly emission¹⁰ of typical floral-odour compounds. Flowers of *O. sphegodes*, like those of many *Ophrys* species, emit only minute quantities of such volatiles^{5,11}. Chemical

Table 1 Electrophysiologically active compounds in cuticle extracts of virgin *Andrena nigroaenea* females and labellum extracts of *Ophrys sphegodes* flowers

Compound	Abundance (%)		Statistics <i>P</i>
	<i>Andrena</i>	<i>Ophrys</i>	
Henicosane	1.6 $\pm$ 0.5	1.8 $\pm$ 0.3	0.151
Docosane	0.6 $\pm$ 0.1	0.5 $\pm$ 0.1	0.909
Tricosane	28.7 $\pm$ 2.4	30.6 $\pm$ 1.8	0.705
Tetracosane	2.0 $\pm$ 0.2	3.1 $\pm$ 0.2	0.002
(<i>Z</i>)-9-pentacosene	3.4 $\pm$ 1.8	0.6 $\pm$ 0.1	0.406
Pentacosane	34.9 $\pm$ 2.2	20.2 $\pm$ 1.3	0.000
Hexacosane	1.6 $\pm$ 0.1	2.1 $\pm$ 0.2	0.016
(<i>Z</i>)-12 + (<i>Z</i>)-11-heptacosene	0.7 $\pm$ 0.3	6.0 $\pm$ 0.8	0.000
(<i>Z</i>)-9-heptacosene	5.1 $\pm$ 1.6	7.6 $\pm$ 1.0	0.112
Heptacosane	11.2 $\pm$ 1.1	11.5 $\pm$ 1.5	0.940
(<i>Z</i>)-12 + (<i>Z</i>)-11-nonacosene	3.7 $\pm$ 1.4	6.7 $\pm$ 1.0	0.010
(<i>Z</i>)-9-nonacosene	6.6 $\pm$ 0.4	9.4 $\pm$ 1.2	0.082

Abundances are given as mean $\pm$ s.e.m. ($n = 10$); *P* values were calculated using the Mann–Whitney *U*-test.

mimicry in the sexually deceptive pollination system therefore also provided an economical way for the plant to ensure its transfer of gametes.

Florian P. Schiestl*, **Manfred Ayasse***,
Hannes F. Paulus*, **Christer Löfstedt†**,
Bill S. Hansson†, **Fernando Ibarra‡**,
Wittko Francke‡

*Institute of Zoology,

Department of Evolutionary Biology,
Althanstrasse 14, A-1090 Vienna, Austria
e-mail (F.P.S.): a8917722@univie.ac.at

†Department of Ecology, Lund University,
S-22362 Lund, Sweden

‡Department of Organic Chemistry,
University of Hamburg,
20146 Hamburg, Germany

1. Pouyanne, A. *Bull. Soc. Hist. Nat. Afr. Noire* **8**, 6–7 (1917).
2. Kullenberg, B. *Studies in Ophrys Pollination* (Almqvist & Wiksells, Uppsala, 1961).
3. Paulus, H. F. & Gack, C. *Isr. J. Bot.* **39**, 43–79 (1990).
4. Linnaeus, C. *Öländska och Gothländska Resa* (Stockholm, 1745).
5. Borg-Karlson, A.-K. *Phytochemistry* **29**, 1359–1387 (1990).
6. Bergström, G. in *Biochemical Aspects of Plant and Animal Coevolution* (ed. Harborne, G.) 207–230 (Academic, New York, 1978).
7. Borg-Karlson, A.-K. & Tengö, J. *J. Chem. Ecol.* **12**, 1927–1941 (1986).
8. Baker, E. A. in *The Plant Cuticle* (eds Cutler, D. F., Alvin, K. L. & Price, C. E.) 139–165 (Academic, London, 1982).
9. Nilsson, L. A. *Trends Ecol. Evol.* **7**, 255–259 (1992).
10. Vogel, S. *Akad. Wiss. Lit. (Mainz) Abh. Math.-Nat. Kl.* **10**, 1–165 (1962).
11. Schiestl, F. P., Ayasse, M., Paulus, H. F., Erdmann, D. & Francke, W. *J. Chem. Ecol.* **23**, 2881–2895 (1997).

Fractal analysis of Pollock's drip paintings

Scientific objectivity proves to be an essential tool for determining the fundamental content of the abstract paintings produced by Jackson Pollock in the late 1940s. Pollock dripped paint from a can onto vast canvases rolled out across the floor of his barn. Although this unorthodox technique has

been recognized as a crucial advancement in the evolution of modern art, the precise quality and significance of the patterns created are controversial. Here we describe an analysis of Pollock's patterns which shows, first, that they are fractal¹, reflecting the fingerprint of nature, and, second, that the fractal dimensions increased during Pollock's career.

To quantify the fractal content of Pollock's drip paintings, such as *Alchemy* (Fig. 1), we used the well-established 'box-counting' method² to calculate the fractal dimension D . We cover the scanned photograph of a Pollock painting with a computer-generated mesh of identical squares. The number of squares $N(L)$ that contain part of the painted pattern is then counted; this is repeated as the size, L , of the squares in the mesh is reduced. The largest size of square is chosen to match the canvas size ($L \approx 2.5$ m) and the smallest is chosen to match the finest paintwork ($L \approx 1$ mm). For fractal behaviour, $N(L)$ scales according to $N(L) \propto L^{-D}$, where $1 < D < 2$. The D values are extracted from the gradient of a graph of $\log N(L)$ plotted against $\log L$. This fractal analysis reveals two distinct D values occurring over the ranges $1 \text{ mm} < L < 5 \text{ cm}$ and $5 \text{ cm} < L < 2.5 \text{ m}$. Our analysis of a film of Pollock while painting shows that the fractal patterns occurring over the lower range are determined by the dripping process, whereas the fractal patterns across the higher range are shaped by his motions around the canvas.

Our analysis shows that Pollock refined his dripping technique: the fractal dimensions increased steadily through the years, from close to 1 in 1943 to 1.72 in 1952. Because D follows such a distinct evolution with time, the fractal analysis could be used as a quantitative, objective technique both to validate and date Pollock's drip paintings. The change in D reflects a dramatic evolution in visual character. His initial drip paintings of 1943 consisted of a single layer

of paint trajectories that occupied only 20% of the 0.35-m^2 canvas area; by 1952 he was painting multiple layers of trajectories that covered over 90% of his 9.96-m^2 canvas. It is important that Pollock introduced his fractals systematically: the initial fractal layer essentially determined D by acting as an anchor layer for the subsequent fractal layers, which then fine-tuned the value of D .

Richard P. Taylor, Adam P. Micolich,
David Jonas

School of Physics, University of New South Wales,
Sydney 2052, Australia
e-mail: rpt@newt.phys.unsw.edu.au

1. Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, New York, 1977).
2. Ott, E. *Chaos in Dynamical Systems* (Cambridge Univ. Press, 1993).

Release from inhibition reveals the visual past

Prolonged viewing of a high-contrast repetitive pattern such as a grating leads to adaptation of the corresponding visual-processing channels¹. We have found that such viewing also leads to the short-term establishment of a subthreshold trace in the brain that can cause a visual illusion of the pattern during rebound from the cross-orientation inhibition^{2–4} that is induced by viewing moving patterns with an orthogonal orientation.

McKay⁵ reported that viewing a stationary pattern of radial lines causes the illusion of fine grains or wavy lines moving in circles. This has been explained as being due to antagonism between pattern and movement channels within an orientation column⁶. We have discovered a new visual after-effect by extending this model to explore antagonism between different orientation columns.

With the subject fixating at the centre, a windmill pattern is presented that slowly rotates at 0.2 Hz in one direction for 5 s, and then in the other direction (Fig. 1a). If this is followed by a pattern orthogonal to the windmill, namely concentric rings, diverging and converging for 5 s each at 2 Hz, viewing a blank screen after the concentric rings causes a vivid after-effect of a stationary windmill for a few seconds. This after-effect is quite different from that seen after viewing the concentric rings alone at 2 Hz, if indeed any after-effect is visible. With a cyclical presentation of the sequence shown in Fig. 1a, a striking windmill-like after-effect is always seen after the concentric rings (bottom left blank in Fig. 1a), but no comparable after-effect is visible immediately after the windmill (top right blank in Fig. 1a). We found that the after-effect could be produced if the rings were delayed by up to 30 s, but not after 60 s.



Figure 1 *Alchemy*, painted by Jackson Pollock in 1947. Drip paintings of this period are characterized by fractal dimensions close to 1.5. Reproduced by permission of ARS, NY and DACS, London, 1999.

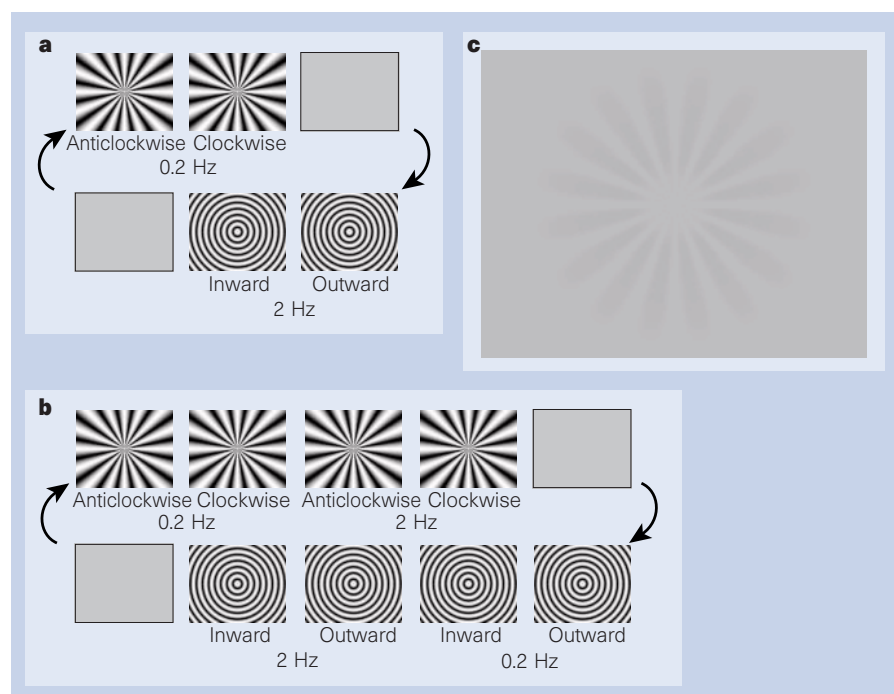


Figure 1 The stimuli used and the after-effect. **a, b**, Two different sequences of images that were used to test for the after-effect. The stimuli were produced with a VSG series 3 stimulus generator (Cambridge Research System, Rochester, UK). The radial patterns rotate clockwise or anticlockwise and the concentric rings move either inwards or outwards. The grey image is the blank screen of the same average luminance (31.3 cd m^{-2}) as the patterns. The screen was placed at a distance of 85 cm. **c**, A simulated enlarged image of the windmill after-effect that most observers see (lower left blank in **a** and **b**). The contrast approximates to the strength of the percept. The Supplementary Information includes a demonstration of the illusions as in **a** and **b**, as well as the script for viewing them on the VSG system.

The after-effect in Fig. 1a resembles the windmill pattern and is not related to the spatial frequency of the rings. If the angular frequency of the windmill changes, so does that of the after-effect. An essential requirement for the after-effect we describe is that the windmill pattern should be moving slowly, at 1 Hz or less, indicating that the stimulation of sustained pattern channels⁷ is crucial to the illusion.

The version shown in Fig. 1b demonstrates the effect more vividly. If the windmill and concentric patterns are both presented at a slow and then at a faster rate (for example, 0.2 Hz and then 2 Hz) with intervening blank screens, and the routine is cycled, after-effects can be seen in each blank period. In the period after the windmill (Fig. 1b, top right), a concentric ring after-effect can be seen; in the period after the concentric rings (bottom left), a windmill after-effect is seen. We presented these patterns to 13 naive subjects, who all reported seeing the illusions. Figure 1c approximates the windmill after-effect that most subjects perceive. It is unlikely that a retinal after-image could produce the illusion because the inducing patterns were not stationary and the after-effects seen were not of the immediately preceding patterns.

Even though the effects could be obtained with simple sine-wave gratings, we

recommend radial patterns to avoid retinal after-images from pursuit eye movements. The after-effect is also stronger with radial patterns, which might be related to the meridional bias reported at many levels of the visual system^{4,8,9}.

If, after the windmill pattern, a moving grating is presented instead of concentric rings, the after-effect is seen in an hourglass fashion, with the radial lines only at orientations orthogonal to the grating. This orientation selectivity, together with our finding that the effects are seen better binocularly than monocularly, suggests that a cortical locus is involved.

We propose the occurrence of two neuronal processes to explain the illusion. First, the stimulation of pattern channels not only leads to a decrease in sensitivity for that pattern¹, but also causes the deposition of a trace lasting for about half a minute, possibly at the synaptic level, akin to some form of short-term plasticity¹⁰. Second, this trace is brought to perceptual threshold by a rebound from inhibition caused by orthogonal orientations.

Georgeson⁶ explained McKay's rays by proposing that there is mutual antagonism between pattern and movement channels within the same orientation column. We suggest that this antagonism between pattern and movement channels also exists between orthogonal orientation columns.

The phenomenon that we have observed might serve as a strategy of the visual system for preventing the spurious excitation of contour detectors. Visual cortical cells, although best stimulated by bars moving perpendicular to the long axis of their receptive fields, often respond to small stimuli moving along this axis¹¹. During eye movements, texture elements in the scene could therefore inappropriately stimulate cells whose optimum orientation is the same as the direction of movement. We propose that this excitation is prevented by inhibition from cells tuned to orthogonal orientations when the latter are activated by moving stimuli.

The transient existence of a stimulus trace is hard to explain. It might be an inevitable consequence of synaptic processes underlying short-term or long-term potentiation. It might also be a way of rapidly facilitating the perception of pre-existing contours after an eye movement.

T. R. Vidyasagar*, **P. Buzás†**,
Z. F. Kisvárdy†, **U. T. Eysel†**

**Psychobiology Laboratory, Division of Psychology, Australian National University, Canberra, ACT 0200, Australia*

†Department of Neurophysiology, Faculty of Medicine, Ruhr-University Bochum, 44780 Bochum, Germany
e-mail: sagar@anu.edu.au

1. Blakemore, C. & Campbell, F. W. *J. Physiol., Lond.* **203**, 237–260 (1969).
2. Creutzfeldt, O. D., Kuhnt, U. & Benevento, L. A. *Exp. Brain Res.* **21**, 251–274 (1974).
3. Sillito, A. M. *J. Physiol., Lond.* **250**, 305–329 (1975).
4. Vidyasagar, T. R., Pei, X. & Volgushev, M. *Trends Neurosci.* **19**, 272–277 (1996).
5. McKay, D. M. *Nature* **180**, 849–850 (1957).
6. Georgeson, M. *Nature* **259**, 413–415 (1976).
7. Kulikowski, J. J. & Tolhurst, D. J. *J. Physiol., Lond.* **232**, 149–162 (1973).
8. Levick, W. R. & Thibos, L. N. *J. Physiol., Lond.* **329**, 243–261 (1982).
9. Shou, T. & Leventhal, A. G. *J. Neurosci.* **9**, 4287–4302 (1989).
10. Fisher, S. A., Fischer, T. M. & Carew, T. J. *Trends Neurosci.* **20**, 170–177 (1997).
11. Wörgötter, F. & Eysel, U. T. *Exp. Brain Res.* **76**, 307–314 (1989).

Supplementary information is available on *Nature's* World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of *Nature*.

Climate change related to egg-laying trends

Analysis of 20 species of UK breeding birds over a 25-year period found a long-term trend towards earlier egg-laying¹. Further studies have correlated such trends with spring temperatures (one species)² or the North Atlantic Oscillation (three species)³. We have studied a data set spanning 57 years and find that laying date is related to temperature or rainfall for 31 of 36 species (86%), and that 53% of species show long-term trends in laying date over time, of which 37% can be statistically accounted

for by changes in climate. These data provide evidence for the large-scale impact of rising temperatures on wildlife. Our analysis of a UKCIP98 national-level climate scenario⁴ predicts that average laying dates will be even earlier for 75% of species by the year 2080.

We analysed 92,828 records gathered from throughout the United Kingdom between 1939 and 1995 by the British Trust for Ornithology's Nest Record Scheme (see Supplementary Information). Annual median laying dates for 36 species, for which we had at least 1,000 records each, were used to measure shifts in the timing of whole nesting seasons. In most cases, the upper and lower tails of the distributions shifted in parallel to the medians, but the latter were less subject to random variation when annual samples were relatively small. Of these 36 species, 19 (53%) exhibited significant long-term trends, most of which were curvilinear, with laying dates becoming later in the 1960s and 1970s, and then earlier through the 1980s and 1990s (Fig. 1). Trends towards earlier laying in recent years are due to increased numbers of early nests and were never due solely to fewer records of late nests.

This data set does not allow us to distinguish first clutches from subsequent ones, or from renests after failure, and species might extend their nesting seasons as a consequence of earlier nesting opportunities. Although this study is therefore statistically conservative, and so might underestimate trends towards earliness (and the effect of weather on laying dates), only the robin *Erithacus rubecula* (which has several broods a year) showed evidence of significant increases in the incidence of early and late nests, using a measure of the width of the two tails of the distribution.

We investigated relationships of median laying date with temperature and rainfall using the Central England Temperature (CET)⁵ and the England and Wales Precipitation (EWP)⁶ records, which are broadly representative of UK weather⁷. We used forward stepwise regression to identify the mean monthly temperature and precipitation variables that accounted for variation in median laying date, weighted by sample size. Weather variables were entered only for months in which laying had been recorded and for the previous month. Significant CET or EWP effects were found for 31 species (86% of the total); the five species without effects were among those with relatively small average annual sample sizes, which may have limited the statistical power. Given the number of associations between month and weather that were tested for significance, some results might be due to chance effects, although more than half the species exhibited weather relationships at $P < 0.001$ and only six species were solely

dependent upon effects at $P > 0.01$ (see Supplementary Information).

Of 19 species with long-term trends in laying date, 17 species (89%) exhibited significant weather effects (Fig. 1). The key weather variables were the temperatures in March and April, which show a strong curvilinear trend since the 1940s, having cooled in the 1960s and 1970s (Fig. 1). When trend variables (year, year² or year³, as appropriate) were included in the weather models for 7 (41%) of these 17 species, trend variables were no longer significant. Thus, variation in CET and EWP was sufficient to account for these long-term trends, describing, on average, 29% more of the variation in laying date than the trend models alone. These seven species (winter wren, *Troglodytes troglodytes*; dun-nock, *Prunella modularis*; blackcap, *Sylvia atricapilla*; willow warbler, *Phylloscopus trochilus*; spotted flycatcher, *Muscicapa striata*; long-tailed tit, *Aegithalos caudatus*; and greenfinch, *Carduelis chloris*) are all widespread in Britain, and it is likely that their records provide the best match with the weather data. For a further ten species, trend variables had additional effects to weather variables, accounting for, on average, 16% more than the weather models alone.

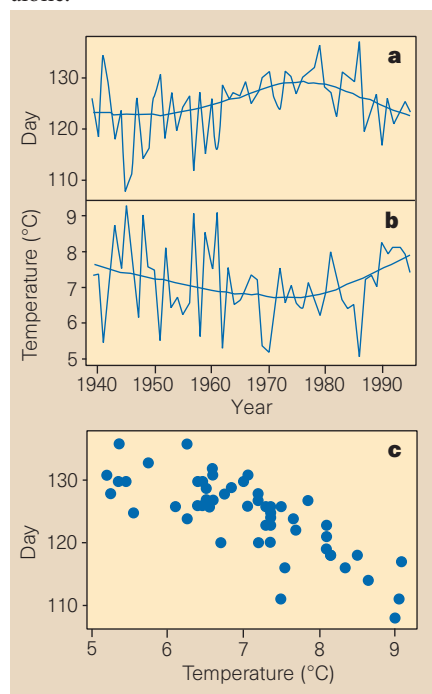


Figure 1 Weather and laying dates of the chaffinch, *Fringilla coelebs*. **a**, Temporal changes in annual median laying date. **b**, Temporal changes in the mean of March and April monthly mean Central England Temperatures (CETs). **c**, Relation between annual median laying date and mean of March and April CETs ($r = -0.76$, $P < 0.001$). Laying date is numbered such that day 110 is 20 April and day 121 is 1 May. The smoothed lines are calculated using a LOWESS (locally weighted scatterplot smoother) method.

These residual trends (and those of the two species without any CET or EWP effects) could be due to the choice of weather variables (for example, the temperature of the soil, rather than the air, might have been more appropriate for some species), or because birds may respond to environmental factors (such as food supply) for which CET and EWP are to some extent surrogates; in addition, the residual trends may be unrelated to climate (for example, large-scale changes in land use might have affected some farmland birds, and magpies may nest earlier because of their rapid expansion into urban areas).

There appear to be no obvious taxonomic, ecological or life-history correlates of the species that are more or less affected by weather. Within each of the groups identified in the Supplementary Information there are, allowing for the size of the groups, mixtures of both residents and migrants, insectivores and granivores, and single- and multibrooded species. The only possible correlate might be body size, as the small number of larger-bodied species (corvids and water birds) appear to show little response to temperature.

We have used the 'medium-high' UKCIP98 climate scenario⁴ for Britain to predict how laying dates for each species may change by 2080, based on the (linear) coefficients of temperature and precipitation. For the 27 species that are predicted to lay earlier, the average advancement is 8 days, ranging up to 18 days (see Supplementary Information). The significance of these findings for nesting success and the subsequent survival of young needs now to be investigated, especially with respect to the possibility of mistimed reproduction in relation to food supplies^{2,8}.

Humphrey Q. P. Crick*,

Timothy H. Sparks†

*British Trust for Ornithology,
The National Centre for Ornithology,
Thetford, Norfolk IP24 2PU, UK

†Institute of Terrestrial Ecology,
Monks Wood, Abbots Ripton, Huntingdon,
Cambridgeshire PE17 2LS, UK
e-mail: h.crick@bto.org

1. Crick, H. Q. P., Dudley, C., Glue, D. E. & Thomson, D. L. *Nature* **388**, 526 (1997).
2. McCleery, R. H. & Perrins, C. M. *Nature* **391**, 30–31 (1998).
3. Forchhammer, M. C., Post, E. & Stenseth, N. C. *Nature* **391**, 29–30 (1998).
4. Hulme, M. & Jenkins, G. J. *Climate Change Scenarios for the UK: Scientific Report* (UKCIP Technical Report No. 1, Climatic Research Unit, Norwich, 1998).
5. Parker, D. E., Legg, T. P. & Folland, C. K. *Int. J. Climatol.* **12**, 317–342 (1992).
6. Jones, P. D. & Conway, D. *Int. J. Climatol.* **17**, 427–438 (1997).
7. Hulme, M. & Jones, P. D. *Environ. Pollut.* **83**, 23–36 (1994).
8. Visser, M. E., van Noordwijk, A. J., Tinbergen, J. M. & Lessells, C. M. *Proc. R. Soc. Lond. B* **265**, 1867–1870 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Captain, art thou sleeping there below?

The Restless Sea

by Robert Kunzig

W. W. Norton: 1999. 336 pp. \$24.95, £16.95

Alan Longhurst

If Francis Drake, “slung between the round-shot, in Nombre de Dios Bay”, had tobogganed down the continental slope into deep water, he would not, as Robert Kunzig pretends, have been crushed “in an instant” to a small fraction of his former volume, like the Styrofoam cups that oceanographers delight in attaching to their instruments for the subsequent bafflement of their spouses. Instead, sewn into sailcloth, he might still be in quite good shape, integrating gently into the dark world of the depths, to judge by the famous apples recovered from the sunken submarine *Alvin*. Bacterial physiology at great pressure runs slowly, except in special places.

Science is a minefield into which even a specialized journalist treads at his peril, and Kunzig’s text slips out of focus now and then. The trick is to keep going, ignoring what doesn’t seem quite right. For the lay reader, this probably won’t matter, because the large fabric is clearly explained; this recounting of recent advances in the ocean sciences presents a lot of fundamental Earth science in easily understood terms.

In ten chapters, some of the jam — but not much of the bread and butter — of oceanography is reviewed. From the motions of tectonic plates and the dynamics of the consequential mid-ocean ridges, we travel to the sulphur-based benthic ecosystems of the rift valleys, the gelatinous plankton of the great depths and the iron-starved algal cells of the Southern Ocean. Finally, we reach the great conveyor belt of the oceanic thermohaline circulation and the possible consequences for global climate of a change in its state. These chapters are sandwiched between accounts of the probable cosmic origins of water on Earth and how it will be dissipated as our Sun becomes a red giant.

Each chapter is a conversational retelling of how American oceanographers solved enigmas over the past 40 years or so. Idiosyncratic personalities and the informality of oceanography in the 1960s and 1970s come across rather well.

As I read, I began to speculate why solutions appeared when they did. Had nobody before the young Hank Stommel thought to apply the conservation of angular momentum to explain the intensification of western boundary currents like the Gulf Stream? If a first-year graduate student with “sulphur bacteria on her brain” had not gone to a seminar on the anatomy of *Riftia*, how long would we have waited to discover the symbiotic sulphur-oxidizing bacteria living

in the trophosome of this abyssal tube worm? And from this, that Earth has two ecosystems, one based on the oxidation of geothermal hydrogen sulphide in abyssal rift valleys, the other based on energy from sunlight. In this case, somebody must surely have spotted quite soon how the exuberant growth of the vent fauna is supported.

And what about iron limitation of phytoplankton in the high nitrate/low chlorophyll areas of the Southern Ocean and parts of the eastern Pacific? On EASTROPAC, the 1960s international expedition, we were already at sea — literally and figuratively — looking for the missing nutrient. But it was John Martin, unwittingly prepared by his mastery of trace-metal assay, who solved the problem in 1986. Had we earlier realized that high nitrate/low chlorophyll areas correspond with places where little terrestrial dust falls on the ocean, the solution might have been found more quickly.

But history is much more complex than Kunzig allows, and although the simple historical introductions to each chapter serve his purpose well enough, they obscure the

true complexity of events. For example, Eric Mills’ *The Sea* (Wiley-Interscience), a rich account of the exploration of deep-sea biology, suggests several places where Kunzig missed the point. But at least the flavour is there for the lay reader. I particularly liked Kunzig’s evocation of the dark world of the plankton deep in the ocean, where relative water motion is negligible and everybody waits for an invisible somebody else to make a move so they can be grabbed.

Kunzig hints at, but doesn’t quite grasp, the speed at which the practice of ocean science is changing. The submersibles once used to explore the ocean depths are unavailable or otherwise employed; research funds are now allocated to coordinated projects with practical aims. Oceanography has entered a new phase, in the sense of Mills — for whom oceanography, unlike mathematics, has no internal dynamic, but instead is driven by external forces. At the end of the nineteenth century, British marine science was turned deliberately away from deep-sea exploration towards coastal biology and fisheries, a policy that has marked it to this



In shallower waters

Australian photographer Roger Steene works with marine scientists to record sea life around the world. Among the 340 colour photographs in his *Coral Seas* (Firefly, \$50) are creatures — from microscopic organisms to a super-mimic octopus — that have not yet been officially

discovered or named. Those that have include, from top, the pink anemonefish, whose body mucus contains chemicals that protect it from its host’s stinging cells; the triggerfish, whose dorsal fins lock upright to wedge it safely into a sleeping space; and the rhinoceros triplefin.

day. Now the Cold War is over, the US Navy is losing interest in basic deep-sea science, and executive empowerment is eroding the initiative of individual oceanographers. Russian institutions are in deep recession, and in Canada there has been a massive diversion of oceanographic resources to the problems of the collapsed fisheries.

Perhaps the next phase of oceanography will see an even greater divergence between the ocean and its coastal regions, between global climate-related studies in the ocean basins and pollution- and fishery-related problems around the edges. The curiosity-driven initiatives that served us so well and revolutionized our understanding of the oceans in the previous phase from 1950 to 1990 will wither. Those of us who participated were indeed fortunate.

But, returning to Kunzig, most writing about the sea, even in the scientific literature, addresses only coastal or continental-shelf phenomena, and I value this book for its almost exclusive attention to the neglected deep oceans. If it is read by enough lay people, perhaps I shan't have to explain so often that, no, oceanographers don't all go to sea wearing red watch caps, and we don't all have scuba tanks slung over our shoulders. □

Alan Longhurst is at Place de l'Eglise, 46160 Cajiarc, France.

More from the depths

Essentials of Oceanography, 6th edn

by Al Trujillo & Hal Thurman

Prentice Hall: 1999. 526 pp.

\$42.99, £22.99 (pbk)

Firing up the power plant in our heads

Evolving Brains

by John Morgan Allman

W. H. Freeman/Scientific American: 1999. 256 pp. \$34.95, £23.95

Bob Martin

Although this book is one of tens in the Scientific American series, it is quite remarkable in presenting a radically novel picture of the evolution of the brain. A major milestone in this field, as John Allman himself notes, was the publication in 1973 of Harry Jerison's monograph *Evolution of the Brain and Intelligence*. That treatise wove many disparate threads into an entirely new tapestry and stimulated much subsequent research. *Evolving Brains* now provides us with another milestone — it is a masterpiece of synthetic presentation and evocative interpretation.

An appreciation of biological continuity can greatly enrich our understanding of the evolution of brains, as Allman elegantly demonstrates. Indeed, he goes back as far as the first unicellular organisms in his search

for guiding principles, and reminds us, for instance, that photoreceptors and olfactory receptors are derived from modified cilia. This illustrates the fact that evolution of the CNS, just like that of any other organ system, has depended heavily on conversion and duplication of pre-existing structures: innovation wedded to an astonishing degree of conservatism. Perhaps the most striking example of conversion is provided by genes with the ancient function of controlling formation of the gut, which now control growth of the forebrain in mammals.

Allman goes on to show that we can learn much about the organization and functioning of the human brain through lessons drawn from evolutionary history. To do so, he considers two contrasting components of the CNS: the ancient serotonergic system of chordates (involving a gene family of serotonin receptors) and the neocortex (a novel structure in mammals).

He draws connections between duplication of genes, duplication and modification of structures in development through corresponding changes in homeotic genes, rhombomeres (repeating segments in the hindbrain), and possible duplication of topographic maps in the neocortex. His use of a power-plant analogy, for the progressive overlaying of new structures on an organ that can never cease functioning, is an example of his ability to create graphic metaphors.

The strength of Allman's review is clearly attributable to his unusual background. He originally studied anthropology at the University of Chicago, during which time he became aware of the special significance of vision in primate evolution and developed an interest in the transfer of the visual image from the retina to the brain. In order to pursue this, he obviously needed to study neurophysiology. His adviser, Clark Howell, suggested that he could continue to be based in Chicago while working at the laboratory of neurophysiology at the University of Wisconsin. There, under the guidance of Clinton Woolsey and Wally Welker, he initiated his primary research on primate visual systems and began his collaboration with Jon Kaas on mapping studies of the visual cortex in owl monkeys.

In 1974, he moved to the California Institute of Technology and has remained there ever since, continuing his research into the primate central nervous system (CNS). Throughout, his training in anthropology has been reflected not only in a broad-based approach to primate evolution, but also in a strong emphasis on biological continuity and historical background. All these factors are combined in the thought-provoking synthesis he presents in *Evolving Brains*.

The comparative approach inevitably looms large in his synthesis. Comparisons show, for example, that the neocortex in all mammals (but not other vertebrates) is a six-

layered structure developed from the roof of the forebrain. Hence, the neocortex is as much a defining characteristic of mammals as are mammary glands, hair or ear ossicles.

In a more restricted example, Allman and his colleagues have shown that, among mammals, primates have a special pattern of visual projection in the optic tectum that eliminates redundancy by representing only one half of the binocular field on each side of the brain. (It was this discovery that first drew my attention to Allman's pioneering work in 1980.) This may be linked with their developing an ability to fix their gaze on objects of interest — primates, in fact, have an apparently unique centre for visual guidance of body movements, called the ventral premotor area. Incidentally, Allman finds no support for John Pettigrew's proposition, in 1986, that fruit-bats (but not other bats) share this special pattern of visual projection with primates and are therefore related to them. As molecular evidence now overwhelmingly indicates that bats are monophyletic, the 'flying-primate hypothesis' must be regarded as severely disabled.

As with Jerison's book, allometric scaling comparisons also play a prominent role. Allman's research group has made notable contributions to this field in recent years. For instance, it was demonstrated that the long-mooted connection between brain size and lifespan is confined to particular brain parts and is (for some reason) less clear in prosimians than in simian primates. Subsequently it was shown that — whereas females typically outlive males in primate species lacking paternal care — males live as long as, or longer than, females in species with extensive paternal care. Allman also rightly emphasizes the central importance of energy relationships, for example, in accounting for the grade shift in relative brain size between cold- and warm-blooded vertebrates.

One weakness in this book is that Allman does not pay much attention to the problem of potential confounding variables — a correlation between brain size and some other feature does not necessarily indicate a direct causal connection. He does not mention the recent debate about potential statistical problems arising from differential relatedness between species in comparative analyses. For instance, leaf-eating monkeys belong to just a few groups of closely related species, so their smaller brains may be traced to just a few ancestral changes. This, however, is just a special case of the general problem that correlation between two variables may exist because both are connected with one or more other variables — numerous additional biological factors as well as phylogenetic relatedness. A case in point is provided by the Aiello/Wheeler 'expensive tissue hypothesis', according to which there was a trade-off between brain size and gut size in primate evolution. Allman integrates this hypothesis

into his synthesis, but it failed to consider potentially confounding variables.

The Scientific American Library series brings together well-produced, accessible reviews for a general readership. This book extends the series in a new direction and breaks new ground in presenting a highly original synthesis. Excellent colour illustrations are a bonus of all the books in this series — the first thing I did with my copy, as with previous examples, was to produce new lecture slides.

Overall, Allman has produced a tightly written, judicious and entertaining account that will surely alter the way our brains think about the evolution of our brains. □

Bob Martin is in the Anthropological Institute, University of Zürich, 8057 Zürich, Switzerland.

The feelgood factor

The Placebo Effect: An Interdisciplinary Exploration

edited by Anne Harrington
Harvard University Press: 1999. 260 pp.
\$19.95, £12.50 (pbk)

John Galloway

The Placebo Effect helps to explain why medicine appears to be some way off relinquishing the certainty of faith for the uncertainty of science. The Latin *placebo* translates as: "I shall be acceptable or pleasing". Traditional medicine views placebos as medicines or procedures that make patients feel better simply by virtue of being given, rather than for what they are. However, scientific respectability has been lent by the current enthusiasm for controlled clinical trials. Placebos are the 'control' against which therapeutic effectiveness is measured.

Despite this, the idea of a placebo suggests a doctor's sleight of hand, a therapeutic illusion where benefit is no more than the patient's faith in their doctor rather than true clinical effectiveness. Is this just a confidence trick, practised by those unwilling to forgo reputation and income by admitting they cannot help, or is it an integral part of good medicine? The jury is still out.

This edited collection of reviews, despite not always being completely digested or integrated, repays reading for the nuggets of insight it gives into health care and its as yet not-so scientific underpinnings. First, does the placebo effect actually exist — or is it, as Elaine and Arthur Shapiro ask in their opening article, "much ado about nothing"? Are placebos distinctive, if nonspecific, in their action, or just part of a spectrum of care and cure? Trying to divide treatments into those that 'really work' and those that do not (even though patients or doctors think they do) raises more questions than it answers. Hence



Take with water (or a pinch of salt): model tablets for the 1936 Chemists' Exhibition in London.

the rise of evidence-based medicine.

In his review, Howard Spiro, a gastroenterologist, uses the example of stomach ulcers to illustrate the problem. For many years, patients were advised to drink milk for immediate pain relief. In the 1970s, this advice was dropped. However, it has been discovered recently that milk stimulates the release of endorphins, giving a 'scientific' reason for its pain-relieving qualities. Will this lead to milk's reinstatement as a treatment? Spiro asks if milk was a placebo then, or an antacid. And, with this new-found knowledge, is it now a placebo, or what?

He points out that cimetidine is now the drug of choice for stomach ulcers. It relieves pain and appears to reduce the size of the ulcer. But the relationship between ulcer size and degree of pain is not clear-cut, nor is the superiority of cimetidine over a placebo in controlled trials — in reducing either pain or the size of the ulcer. In the two-week trial period, half of the ulcers of people using cimetidine had healed, compared with a third for those on the placebo.

The designers of trials believe that the feeling of hopeful expectation induced by even the semblance of treatment is likely to influence how patients react, or at least how those carrying out the trial perceive they react. Otherwise, the placebo arm of the trial would have no point except for saving money. This feature, *de rigueur* for respectable clinical-trial design, is catch-22 for placebos, as it appears to rule out any scientific assessment of the effectiveness of the placebo itself. On the other hand, patients entered in trials do better than those not entered, all else being equal.

In the book's final article, David Morris, the associate editor of the journal *Literature and Medicine*, emphasizes that illness and disease are not always the same, although the one may tell you that you have the other. Disease, its causes and cures, are probably best

viewed as part of biology, with its emphasis on structure and the adaptation of structure to function. But illness may be better understood within frameworks of social anthropology and behavioural psychology.

Literature makes some contribution in this area. The central character in Pat Barker's *Regeneration Trilogy* (Penguin) is the eminent historical figure Dr W. H. R. Rivers, a pioneering neurologist and social anthropologist. During his time on a South Pacific island, Rivers saw how different the concepts of illness and treatment could be in other cultures. When he was injured after his boat foundered, the need to save the boat inspired him to such desperate action that, until the boat was safe, he was unaware of his pain and disability. Thus, under some circumstances, the mind can disregard the pain of injury or disease.

The complementary point is made in *The Placebo Effect*: that people who have lost limbs may continue to feel the pain in the lost part. A strong case is made for an integrated, bio-social view of illness. One implication is that the biological side of medicine may be overplayed at the expense of the social and cultural side. In the United Kingdom, it is possible that too much weight is attached to 'scientific' qualifications for entry to medical courses.

If the placebo effect can be demonstrated scientifically and its underlying mechanisms discovered, it brings together very different medical philosophies, whether science-based, complementary or even magical. This would be something of a paradox. No matter how unscientific the style of practitioners, faith in them would become very much part of science. Is a faith in doctors that is explained by science and, as Robert Brown-ing wrote, "diversified by doubt", better for you as a patient or worse? □

John Galloway is in the Dental Team Studies Unit, Eastman Dental Hospital, London WC1X 8LD, UK.

Forced smile

Francis Crick has pointed out an error in Walter Gratzer's review of Freeman Dyson's *The Sun, the Genome and the Internet* (*Nature* 398, 770-771; 1999). The quotation in the penultimate paragraph was by Oliver Edwards, an acquaintance of Samuel Johnson's, as recorded in James Boswell's *The Life of Dr Johnson*: "You are a philosopher, Dr Johnson. I have tried too in my time to be a philosopher, but, I don't know how, cheerfulness was always breaking in."

From entropy to Duino to Pavia

The correct address for G. F. Bignami, who reviewed Carlo Cercignani's *Ludwig Boltzmann: The Man Who Trusted Atoms* (*Nature* 399, 32-33; 1999), is the Italian Space Agency, Via di Villa Patrizi 13, 00161 Rome, and the University of Pavia.

Science in culture

Cycling the cosmos

Prints of "The Four Seasons"
Martin Kemp

Any mature culture is sustained by intricate systems of belief in which ideas and practices are mutually self-reinforcing. For much of Western culture, from Greek antiquity to the Enlightenment, understanding of the nature of life in the greater scheme of things was founded upon the profoundly cyclical behaviour of the Universe — the gyrations of the heavenly bodies, the phases of the Moon, the passing of the seasons, the times of day.

On this mortal coil the four ages of man acted out their endless succession. This elemental cycle of all natural things operated according to a periodic base of four.

A remarkable and apparently unique set of prints of "The Four Seasons" in the History of Medicine Collections at Duke University in North Carolina is unrivalled as a compact yet comprehensive visual encyclopaedia of the governance of human existence.

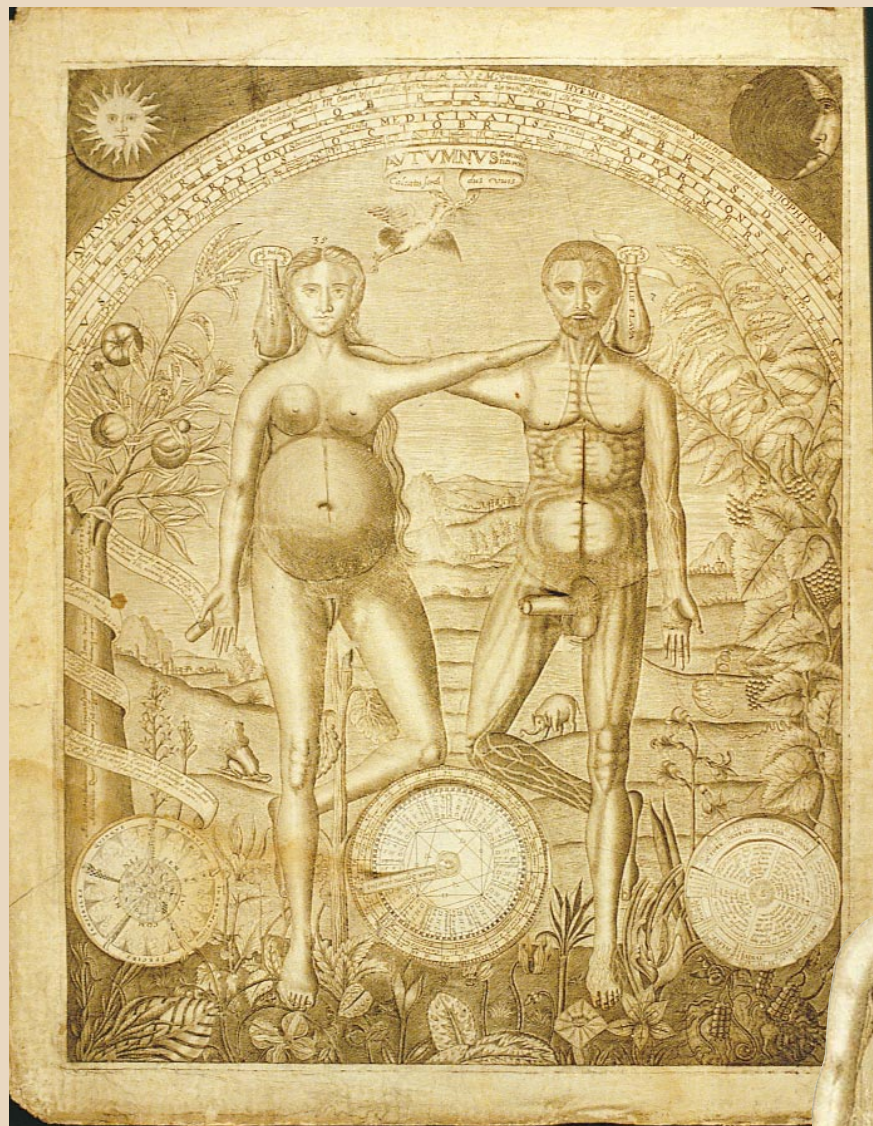
Dating from around 1600 and of northern European or British origin, the prints are of a complexity that defies brief summary — not least because each print is overlaid by strata of superimposed flaps which progressively disclose deeper layers of knowledge, while rotating paper "volvelles" can be used to plot temporal conjunctions of signs.

God's most divine creation, a progressively ageing Adam and Eve, are anatomized under an astrological arch, flanked by the vitalizing Tree of Life and the dangerous Tree of Knowledge. The busy microcosms of animals and plants within the body of the Earth provide the background against which the two archetypal humans pursue their life cycles, accompanied by the remorselessly ticking clock of cosmological time.

In "Autumn", the third in the series, the rich harvest of the implanting of man's seed is made manifest by the woman's swollen belly, which hinges open to reveal the homunculus within.

The Hippocratic mottoes on the ribbons locate the prints within the orbit of the physician's art, which was seen as founded on higher philosophical wisdom. All maladies were treated within the context of humoral medicine, in which the body was governed by four humours: blood, yellow bile, phlegm and black bile. A predominance of one humour would result respectively in the individual manifesting a sanguine, choleric, phlegmatic or melancholic temperament. The humours themselves obeyed the greater governance of the four elements of air, fire, water and earth, and their coupled qualities, moist-hot, hot-dry, cold-moist and dry-cold.

Looking back, it may be difficult to comprehend how a system that appears so erroneous held such sustained sway. But, once



The fullness of time: "Autumn" from the engravings of "The Four Seasons". This is one of a remarkable series that can be seen in "The Physician's Art: Representations of Art and Medicine From Four North Carolina Collections" at Duke University Museum of Art, Durham, North Carolina, from 4 November 1999 to 16 January 2000.

we understand how deeply the parts of the mutually supporting doctrines were embedded in sophisticated patterns of institutional belief and practice, we can better appreciate how each part continued to bear witness to the validity of the others and to the essential rightness of the whole.

Faith in the rule of the planets over our lives, to take just one component, was not unreasonable, given the striking evidence of the Moon's ability to shift great bodies of tidal water, and the cyclical governance evident in all aspects of the natural year.

In societies that were still dominated by the vagaries of agrarian cycles, where the course of events was determined by remote agencies, it was quite rational to believe that some hugely complex clock was at work. There was no reason why humans should prove to be an exception to the rule.

It seems to me that the breakdown in the complex machinery of the set of beliefs in the Four Seasons is profoundly associated with the growing dominance, in the eighteenth century, of the urban environment. This brought an increasing reliance upon technology and the forging of a lifestyle based upon the conviction that scientific and technological progress would accord us mastery over nature.

We are the heirs to this world-picture of human manipulative potential. But there is no guarantee that our systems of interlocking beliefs will not appear as strange to later cultures as those in "The Four Seasons" now appear to us.

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica

J. R. Petit*, J. Jouzel†, D. Raynaud*, N. I. Barkov‡, J.-M. Barnola*, I. Basile*, M. Bender§, J. Chappellaz*, M. Davis||, G. Delaygue†, M. Delmotte*, V. M. Kotlyakov¶, M. Legrand*, V. Y. Lipenkov‡, C. Lorius*, L. Pépin*, C. Ritz*, E. Saltzman|| & M. Stievenard†

* Laboratoire de Glaciologie et Géophysique de l'Environnement, CNRS, BP96, 38402, Saint Martin d'Hères Cedex, France

† Laboratoire des Sciences du Climat et de l'Environnement (UMR CEA/CNRS 1572), L'Orme des Merisiers, Bât. 709, CEA Saclay, 91191 Gif-sur-Yvette Cedex, France

‡ Arctic and Antarctic Research Institute, Beringa Street 38, 199397, St Petersburg, Russia

§ Department of Geosciences, Princeton University, Princeton, New Jersey 08544-1003, USA

|| Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149, USA

¶ Institute of Geography, Staromonetny, per 29, 109017, Moscow, Russia

The recent completion of drilling at Vostok station in East Antarctica has allowed the extension of the ice record of atmospheric composition and climate to the past four glacial–interglacial cycles. The succession of changes through each climate cycle and termination was similar, and atmospheric and climate properties oscillated between stable bounds. Interglacial periods differed in temporal evolution and duration. Atmospheric concentrations of carbon dioxide and methane correlate well with Antarctic air-temperature throughout the record. Present-day atmospheric burdens of these two important greenhouse gases seem to have been unprecedented during the past 420,000 years.

The late Quaternary period (the past one million years) is punctuated by a series of large glacial–interglacial changes with cycles that last about 100,000 years (ref. 1). Glacial–interglacial climate changes are documented by complementary climate records^{1,2} largely derived from deep sea sediments, continental deposits of flora, fauna and loess, and ice cores. These studies have documented the wide range of climate variability on Earth. They have shown that much of the variability occurs with periodicities corresponding to that of the precession, obliquity and eccentricity of the Earth's orbit^{1,3}. But understanding how the climate system responds to this initial orbital forcing is still an important issue in palaeoclimatology, in particular for the generally strong ~100,000-year (100-kyr) cycle.

Ice cores give access to palaeoclimate series that includes local temperature and precipitation rate, moisture source conditions, wind strength and aerosol fluxes of marine, volcanic, terrestrial, cosmogenic and anthropogenic origin. They are also unique with their entrapped air inclusions in providing direct records of past changes in atmospheric trace-gas composition. The ice-drilling project undertaken in the framework of a long-term collaboration between Russia, the United States and France at the Russian Vostok station in East Antarctica (78°S, 106°E, elevation 3,488 m, mean temperature –55 °C) has already provided a wealth of such information for the past two glacial–interglacial cycles^{4–13}. Glacial periods in Antarctica are characterized by much colder temperatures, reduced precipitation and more vigorous large-scale atmospheric circulation. There is a close correlation between Antarctic temperature and atmospheric concentrations of CO₂ and CH₄ (refs 5, 9). This discovery suggests that greenhouse gases are important as amplifiers of the initial orbital forcing and may have significantly contributed to the glacial–interglacial changes^{14–16}. The Vostok ice cores were also used to infer an empirical estimate of the sensitivity of global climate to future anthropogenic increases of greenhouse-gas concentrations¹⁵.

The recent completion of the ice-core drilling at Vostok allows us to considerably extend the ice-core record of climate properties at this site. In January 1998, the Vostok project yielded the deepest ice

core ever recovered, reaching a depth of 3,623 m (ref. 17). Drilling then stopped ~120 m above the surface of the Vostok lake, a deep subglacial lake which extends below the ice sheet over a large area¹⁸, in order to avoid any risk that drilling fluid would contaminate the lake water. Preliminary data¹⁷ indicated that the Vostok ice-core record extended through four climate cycles, with ice slightly older than 400 kyr at a depth of 3,310 m, thus spanning a period comparable to that covered by numerous oceanic¹ and continental² records.

Here we present a series of detailed Vostok records covering this ~400-kyr period. We show that the main features of the more recent Vostok climate cycle resemble those observed in earlier cycles. In particular, we confirm the strong correlation between atmospheric greenhouse-gas concentrations and Antarctic temperature, as well as the strong imprint of obliquity and precession in most of the climate time series. Our records reveal both similarities and differences between the successive interglacial periods. They suggest the lead of Antarctic air temperature, and of atmospheric greenhouse-gas concentrations, with respect to global ice volume and Greenland air-temperature changes during glacial terminations.

The ice record

The data are shown in Figs 1, 2 and 3 (see Supplementary Information for the numerical data). They include the deuterium content of the ice (δD_{ice} , a proxy of local temperature change), the dust content (desert aerosols), the concentration of sodium (marine aerosol), and from the entrapped air the greenhouse gases CO₂ and CH₄, and the $\delta^{18}O$ of O₂ (hereafter $\delta^{18}O_{atm}$) which reflects changes in global ice volume and in the hydrological cycle¹⁹. (δD and $\delta^{18}O$ are defined in the legends to Figs 1 and 2, respectively.) All these measurements have been performed using methods previously described except for slight modifications (see figure legends).

The detailed record of δD_{ice} (Fig. 1) confirms the main features of the third and fourth climate cycles previously illustrated by the coarse-resolution record¹⁷. However, a sudden decrease from interglacial-like to glacial-like values, rapidly followed by an abrupt return to interglacial-like values, occurs between 3,320 and 3,330 m.

In addition, a transition from low to high CO_2 and CH_4 values (not shown) occurs at exactly the same depth. In undisturbed ice, the transition in atmospheric composition would be found a few metres lower (due to the difference between the age of the ice and the age of the gas²⁰). Also, three volcanic ash layers, just a few centimetres apart but inclined in opposite directions, have been observed—10 m

above this δD excursion (3,311 m). Similar inclined layers were observed in the deepest part of the GRIP and GISP2 ice cores from central Greenland, where they are believed to be associated with ice flow disturbances. Vostok climate records are thus probably disturbed below these ash layers, whereas none of the six records show any indication of disturbances above this level. We therefore limit

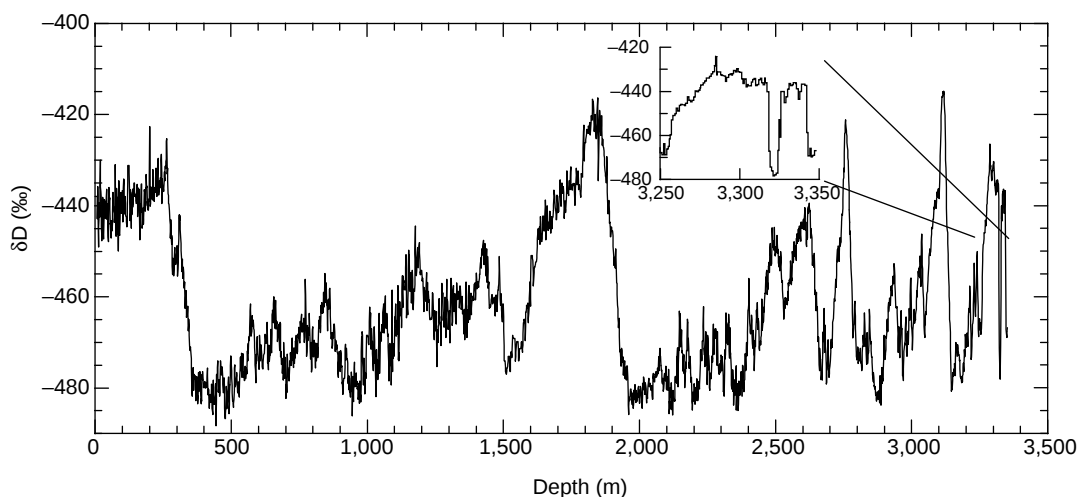


Figure 1 The deuterium record. Deuterium content as a function of depth, expressed as δD (in ‰ with respect to Standard Mean Ocean Water, SMOW). This record combines data available down to 2,755 m (ref. 13) and new measurements performed on core 5G (continuous 1-m ice increments) from 2,755 m to 3,350 m.

Measurement accuracy (1σ) is better than 1‰. Inset, the detailed deuterium profile for the lowest part of the record showing a δD excursion between 3,320 and 3,330 m. $\delta\text{D}_{\text{ice}}(\text{in } \text{‰}) = [(D/H)_{\text{sample}}/(D/H)_{\text{SMOW}} - 1] \times 1,000$.

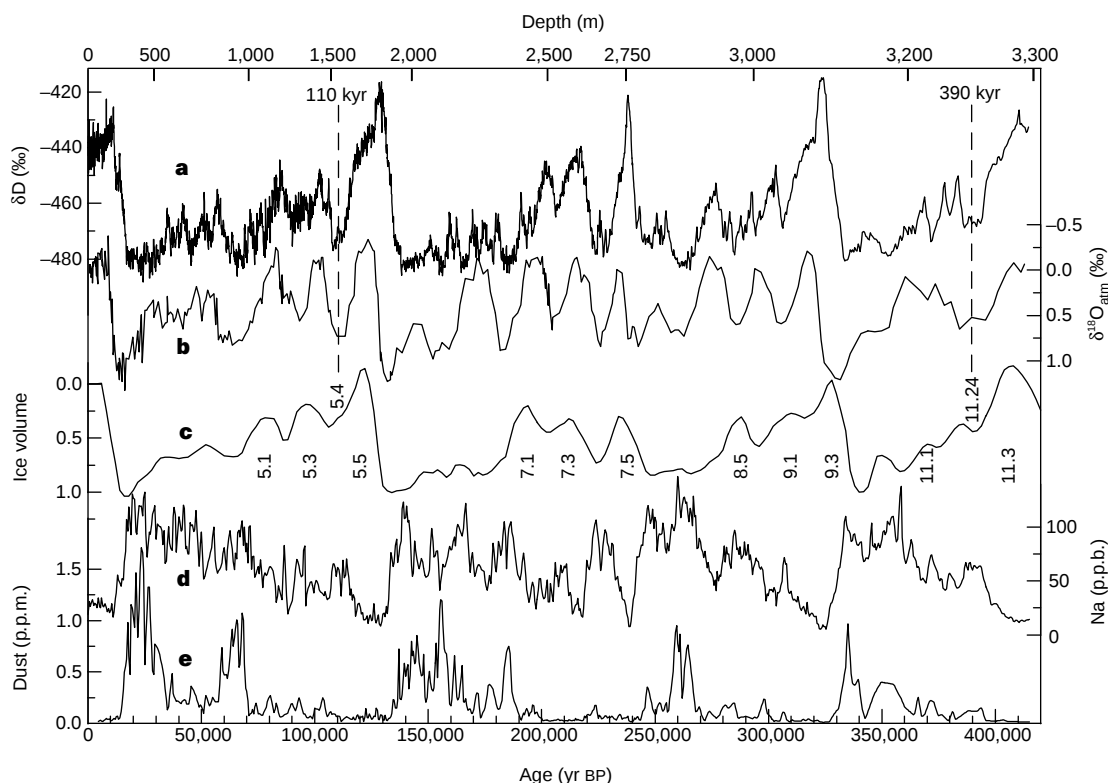


Figure 2 Vostok time series and ice volume. Time series (GT4 timescale for ice on the lower axis, with indication of corresponding depths on the top axis and indication of the two fixed points at 110 and 390 kyr) of: **a**, deuterium profile (from Fig. 1); **b**, $\delta^{18}\text{O}_{\text{atm}}$ profile obtained combining published data^{11,13,30} and 81 new measurements performed below 2,760 m. The age of the gas is calculated as described in ref. 20; **c**, seawater $\delta^{18}\text{O}$ (ice volume proxy) and marine isotope stages adapted from Bassinot *et al.*²⁶; **d**, sodium profile obtained by combination

of published and new measurements (performed both at LGGE and RSMAS) with a mean sampling interval of 3–4 m (ng g^{-1} or p.p.b.); and **e**, dust profile (volume of particles measured using a Coulter counter) combining published data^{10,13} and extended below 2,760 m, every 4 m on the average (concentrations are expressed in $\mu\text{g g}^{-1}$ or p.p.m. assuming that Antarctic dust has a density of $2,500 \text{ kg m}^{-3}$). $\delta^{18}\text{O}_{\text{atm}}(\text{in } \text{‰}) = [(^{18}\text{O}/^{16}\text{O})_{\text{sample}}/(^{18}\text{O}/^{16}\text{O})_{\text{standard}} - 1] \times 1,000$; standard is modern air composition.

the discussion of our new data sets to the upper 3,310 m of the ice core, that is, down to the interglacial corresponding to marine stage 11.3.

Lorius *et al.*⁴ established a glaciological timescale for the first climate cycle of Vostok by combining an ice-flow model and an ice-accumulation model. This model was extended and modified in several studies^{12,13}. The glaciological timescale provides a chronology based on physics, which makes no assumption about climate forcings or climate correlation except for one or two adopted control ages. Here, we further extend the Extended Glaciological Timescale (EGT) of Jouzel *et al.*¹² to derive GT4, which we adopt as our primary chronology (see Box 1). GT4 provides an age of 423 kyr at a depth of 3,310 m.

Climate and atmospheric trends

Temperature. As a result of fractionation processes, the isotopic content of snow in East Antarctica (δD or $\delta^{18}O$) is linearly related to the temperature above the inversion level, T_i , where precipitation forms, and also to the surface temperature of the precipitation site, T_s (with $\Delta T_i = 0.67\Delta T_s$, see ref. 6). We calculate temperature changes from the present temperature at the atmospheric level as $\Delta T_i = (\Delta\delta D_{ice} - 8\Delta\delta^{18}O_{sw})/9$, where $\Delta\delta^{18}O_{sw}$ is the globally averaged change from today's value of seawater $\delta^{18}O$, and 9‰ per °C is the spatial isotope/temperature gradient derived from deuterium data in this sector of East Antarctica²¹. We applied the above relationship to calculate ΔT_s . This approach underestimates ΔT_s by a factor of ~ 2 in Greenland²² and, possibly, by up to 50% in Antarctica²³. However, recent model results suggest that any underestimation of temperature changes from this equation is small for Antarctica^{24,25}.

To calculate ΔT_i from δD , we need to adopt a curve for the change in the isotopic composition of sea water versus time and correlate it with Vostok. We use the stacked $\delta^{18}O_{sw}$ record of Bassinot *et al.*²⁶, scaled with respect to the V19-30 marine sediment record over their common part that covers the past 340 kyr (ref. 27) (Fig. 2). To avoid distortions in the calculation of ΔT_i linked with dating uncertainties, we correlate the records by performing a peak to peak adjustment between the ice and ocean isotopic records. The $\delta^{18}O_{sw}$ correction corresponds to a maximum ΔT_i correction of $\sim 1^\circ C$ and associated uncertainties are therefore small. We do not attempt to correct ΔT_i either for the change of the altitude of the ice sheet or for the origin of the ice upstream of Vostok¹³; these terms are very poorly known and, in any case, are also small ($< 1^\circ C$).

The overall amplitude of the glacial–interglacial temperature change is $\sim 8^\circ C$ for ΔT_i (inversion level) and $\sim 12^\circ C$ for ΔT_s , the temperature at the surface (Fig. 3). Broad features of this record are thought to be of large geographical significance (Antarctica and part of the Southern Hemisphere), at least qualitatively. When examined in detail, however, the Vostok record may differ from coastal²⁸ sites in East Antarctica and perhaps from West Antarctica as well.

Jouzel *et al.*¹³ noted that temperature variations estimated from deuterium were similar for the last two glacial periods. The third and fourth climate cycles are of shorter duration than the first two cycles in the Vostok record. The same is true in the deep-sea record, where the third and fourth cycles span four precessional cycles rather than five as for the last two cycles (Fig. 3). Despite this difference, one observes, for all four climate cycles, the same ‘sawtooth’ sequence of a warm interglacial (stages 11.3, 9.3, 7.5 and 5.5), followed by increasingly colder interstadial events, and ending with a rapid return towards the following interglacial. The

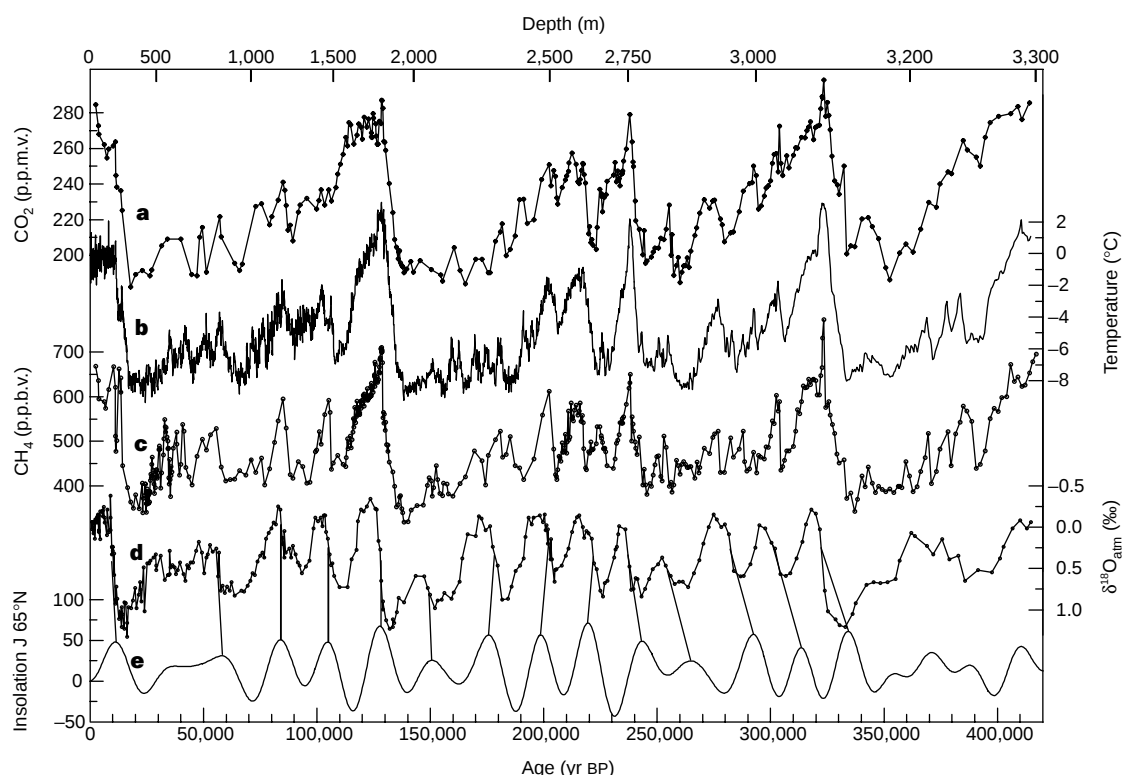


Figure 3 Vostok time series and insolation. Series with respect to time (GT4 timescale for ice on the lower axis, with indication of corresponding depths on the top axis) of: **a**, CO_2 ; **b**, isotopic temperature of the atmosphere (see text); **c**, CH_4 ; **d**, $\delta^{18}O_{atm}$; and **e**, mid-June insolation at $65^\circ N$ (in $W m^{-2}$) (ref. 3). CO_2 and CH_4 measurements have been performed using the methods and analytical procedures previously described^{5,9}. However, the CO_2 measuring system has been slightly modified in order to increase the sensitivity of the CO_2 detection. The

thermal conductivity chromatographic detector has been replaced by a flame ionization detector which measures CO_2 after its transformation into CH_4 . The mean resolution of the CO_2 (CH_4) profile is about 1,500 (950) years. It goes up to about 6,000 years for CO_2 in the fractured zones and in the bottom part of the record, whereas the CH_4 time resolution ranges between a few tens of years to 4,500 years. The overall accuracy for CH_4 and CO_2 measurements are ± 20 p.p.b.v. and 2–3 p.p.m.v., respectively. No gravitational correction has been applied.

coolest part of each glacial period occurs just before the glacial termination, except for the third cycle. This may reflect the fact that the June 65°N insolation minimum preceding this transition (255 kyr ago) has higher insolation than the previous one (280 kyr ago), unlike the three other glacial periods. Nonetheless, minimum

Box 1 The Vostok glaciological timescale

We use three basic assumptions¹² to derive our glaciological timescale (GT4): (1) the accumulation rate has in the past varied in proportion to the derivative of the water vapour saturation pressure with respect to temperature at the level where precipitation forms (see section on the isotope temperature record), (2) at any given time the accumulation between Vostok and Dome B (upstream of Vostok) varies linearly with distance along the line connecting those two sites, and (3) the Vostok ice at 1,534 m corresponds to marine stage 5.4 (110 kyr) and ice at 3,254 m corresponds to stage 11.2.4 (390 kyr).

Calculation of the strain-induced thinning of annual layers is now performed accounting for the existence of the subglacial Vostok lake. Indeed, running the ice-flow model⁴⁸ with no melting and no basal sliding as done for EGT¹² leads to an age >1,000 kyr for the deepest level we consider here (3,310 m), which is much too old. Instead, we now allow for moderate melting and sliding. These processes diminish thinning for the lower part of the core and provide younger chronologies. We ran this age model⁴⁸ over a large range of values of the model parameters (present-day accumulation at Vostok, A , melting rate, M , and fraction of horizontal velocity due to base sliding, S) with this aim of matching the assumed ages at 1,534 and 3,254 m. This goal was first achieved (ages of 110 and 392 kyr) with $A = 1.96 \text{ g cm}^{-2} \text{ yr}^{-1}$, and M and S equal respectively to 0.4 mm yr^{-1} and 0.7 for the region 60 km around Vostok where the base is supposed to reach the melting point (we set $M = 0$ and $S = 0$ elsewhere). These values are in good agreement with observations for A ($2.00 \pm 0.04 \text{ g cm}^{-2} \text{ yr}^{-1}$ over the past 200 yr) and correspond to a reasonable set of parameters for M and S . We adopt this glaciological timescale (GT4), which gives an age of 423 kyr at 3,310 m, without further tuning (Fig. 2). GT4 never differs by more than 2 kyr from EGT over the last climate cycle and, in qualitative agreement with recent results⁴⁹, makes termination I slightly older (by ~700 yr). We note that it provides a reasonable age for stage 7.5 (238 kyr) whereas Jouzel *et al.*¹³ had to modify EGT for the second climate cycle by increasing the accumulation by 12% for ages older than 110 kyr. GT4 never differs by more than 4 kyr from the orbitally tuned timescale of Waelbroeck *et al.*⁵⁰ (defined back to 225 kyr), which is within the estimated uncertainty of this latter timescale. Overall, we have good arguments^{11,50–52} to claim that the accuracy of GT4 should be better than ± 5 kyr for the past 110 kyr.

The strong relationship between $\delta^{18}\text{O}_{\text{atm}}$ and mid-June 65°N insolation changes (see text and Fig. 3) enables us to further evaluate the overall quality of GT4. We can use each well-marked transition from high to low $\delta^{18}\text{O}_{\text{atm}}$ to define a 'control point' giving an orbitally tuned age. The mid-point of the last $\delta^{18}\text{O}_{\text{atm}}$ transition (~10 kyr ago) has nearly the same age as the insolation maximum (11 kyr). We assume that this correspondence also holds for earlier insolation maxima. The resulting control points (Fig. 3 and Table 1) are easy to define for the period over which the precessional cycle is well imprinted in 65°N insolation (approximately between 60 and 340 kyr) but not during stages 2 and 10 where insolation changes are small. The agreement between the $\delta^{18}\text{O}_{\text{atm}}$ control points and GT4 is remarkably good given the simple assumptions of both approaches. This conclusion stands despite the fact that we do not understand controls on $\delta^{18}\text{O}_{\text{atm}}$ sufficiently well enough to know about the stability of its phase with respect to insolation. We assume that the change in phase does not exceed ± 6 kyr (1/4 of a precessional period).

We conclude that accuracy of GT4 is always better than ± 15 kyr, better than ± 10 kyr for most of the record, and better than ± 5 kyr for the last 110 kyr. This timescale is quite adequate for the discussions here which focus on the climatic information contained in the Vostok records themselves.

temperatures are remarkably similar, within 1°C, for the four climate cycles. The new data confirm that the warmest temperature at stage 7.5 was slightly warmer than the Holocene¹³, and show that stage 9.3 (where the highest deuterium value, -414.8‰, is found) was at least as warm as stage 5.5. That part of stage 11.3, which is present in Vostok, does not correspond to a particularly warm climate as suggested for this period by deep-sea sediment records²⁹. As noted above, however, the Vostok records are probably disturbed below 3,310 m, and we may not have sampled the warmest ice of this interglacial. In general, climate cycles are more uniform at Vostok than in deep-sea core records¹. The climate record makes it unlikely that the West Antarctic ice sheet collapsed during the past 420 kyr (or at least shows a marked insensitivity of the central part of East Antarctica and its climate to such a disintegration).

The power spectrum of ΔT_1 (Fig. 4) shows a large concentration of variance (37%) in the 100-kyr band along with a significant concentration (23%) in the obliquity band (peak at 41 kyr). This strong obliquity component is roughly in phase with the annual insolation at the Vostok site^{4,6,15}. The variability of annual insolation at 78°S is relatively large, 7% (ref. 3). This supports the notion that annual insolation changes in high southern latitudes influence Vostok temperature¹⁵. These changes may, in particular, contribute to the initiation of Antarctic warming during major terminations, which (as we show below) herald the start of deglaciation.

There is little variance (11%) in ΔT_1 around precessional periodicities (23 and 19 kyr). In this band, the position of the spectral peaks is affected by uncertainties in the timescale. To illustrate this point, we carried out, as a sensitivity test, a spectral analysis using the control points provided by the $\delta^{18}\text{O}_{\text{atm}}$ record (see Table 1). The position and strength of the 100- and 40-kyr-spectral peaks are unaffected, whereas the power spectrum is significantly modified for periodicities lower than 30 kyr.

Insolation. $\delta^{18}\text{O}_{\text{atm}}$ strongly depends on climate and related properties, which reflect the direct or indirect influence of insolation¹⁹. As a result, there is a striking resemblance between $\delta^{18}\text{O}_{\text{atm}}$ and mid-June insolation at 65°N for the entire Vostok record (Fig. 3). This provides information on the validity of our glaciological timescale

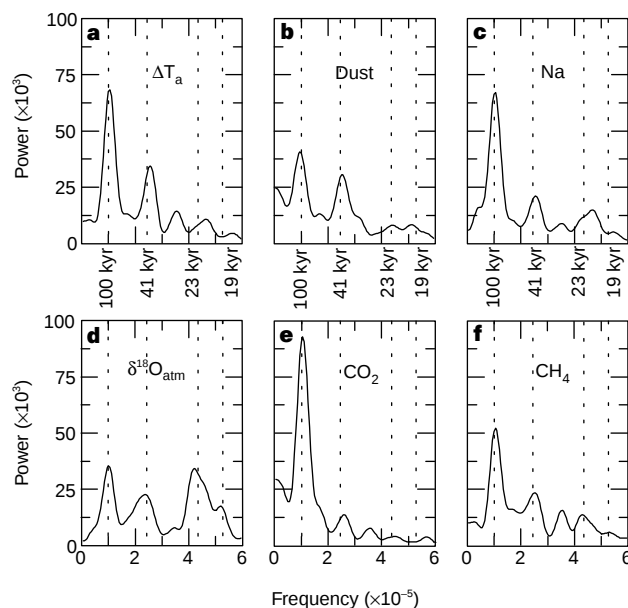


Figure 4 Spectral properties of the Vostok time series. Frequency distribution (in cycles yr^{-1}) of the normalized variance power spectrum (arbitrary units). Spectral analysis was done using the Blackman-Tukey method (calculations were performed with the Analyseries software⁴⁷): **a**, isotopic temperature; **b**, dust; **c**, sodium; **d**, $\delta^{18}\text{O}_{\text{atm}}$; **e**, CO_2 ; and **f**, CH_4 . Vertical lines correspond to periodicities of 100, 41, 23 and 19 kyr.

(see Box 1). The precessional frequencies, which do not account for much variance in ΔT_i , are strongly imprinted in the $\delta^{18}\text{O}_{\text{atm}}$ record (36% of the variance in this band, Fig. 4). In addition, the remarkable agreement observed back to stage 7.5 between the amplitude of the filtered components of the mid-June insolation at 65°N and $\delta^{18}\text{O}_{\text{atm}}$ in the precessional band¹³ holds true over the last four climate cycles (not shown). As suggested by the high variance of $\delta^{18}\text{O}_{\text{atm}}$ in the precessional band, this orbital frequency is also reflected in the Dole effect, the difference between $\delta^{18}\text{O}_{\text{atm}}$ and $\delta^{18}\text{O}_{\text{sw}}$, confirming results obtained on the last two climate cycles^{19,30}.

Aerosols. Figure 2 shows records of aerosols of different origins. The sodium record represents mainly sea-salt aerosol entrained from the ocean surface, whereas the dust record corresponds to the small size fraction ($\sim 2\ \mu\text{m}$) of the aerosol produced by the continent. The extension of the Vostok record confirms much higher fallout during cold glacial periods than during interglacials. Concentrations range up to $120\ \text{ng g}^{-1}$, that is, 3 to 4 times the Holocene value, for sea-salt. For dust, they rise from about $50\ \text{ng g}^{-1}$ during interglacials to $1,000\text{--}2,000\ \text{ng g}^{-1}$ during cold stages 2, 4, 6, 8 and 10.

The sodium concentration is closely anti-correlated with isotopic temperature ($r^2 = -0.70$ over the past 420 kyr). The power spectrum of the sodium concentration, like that of ΔT_i , shows periodicities around 100, 40 and 20 kyr (Fig. 4). Conditions prevailing during the present-day austral winter could help explain the observed glacial/interglacial changes in sodium. The seasonal increase of marine aerosol observed in the atmosphere and snow at the South Pole in September³¹ corresponds to the maximum extent of sea ice; this is because the more distant source effect is compensated by the greater cyclonic activity, and by the more efficient zonal and meridional atmospheric circulation probably driven by the steeper meridional (ocean–Antarctica) temperature gradient. These modern winter conditions may be an analogue for glacial climates, supporting the apparent close anti-correlation between sodium concentration and temperature at Vostok.

The extension of the Vostok dust record confirms that continental aridity, dust mobilization and transport are more prevalent during glacial climates, as also reflected globally in many dust records (see ref. 10 and references therein). The presence of larger particles in the Vostok record, at least during the Last Glacial Maximum¹⁰, indicates that the atmospheric circulation at high southern latitudes was more turbulent at that time. Lower atmospheric moisture content and reduced hydrological fluxes may also have contributed significantly (that is, one order of magnitude³²) to the very large increases of dust fallout during full glacial periods because of a lower aerosol-removal efficiency.

Unlike sodium concentration, the dust record is not well correlated with temperature (see below) and shows large concentrations of variance in the 100- and 41-kyr spectral bands (Fig. 4). The

Vostok dust record is, in this respect, similar to the tropical Atlantic dust record of de Menocal³³ who attributes these spectral characteristics to the progressive glaciation of the Northern Hemisphere and the greater involvement of the deep ocean circulation. We suggest that there also may be some link between the Vostok dust record and deep ocean circulation through the extension of sea ice in the South Atlantic Ocean, itself thought to be coeval with a reduced deep ocean circulation³⁴. Our suggestion is based on the fact that the dust source for the East Antarctic plateau appears to be South America, most likely the Patagonian plain, during all climate states of the past 420 kyr (refs 35, 36). The extension of sea ice in the South Atlantic during glacial times greatly affects South American climate, with a more northerly position of the polar front and the belt of Westerlies pushed northward over the Andes. This should lead, in these mountainous areas, to intense glacial and fluvial erosion, and to colder and drier climate with extensive dust mobilization (as evidenced by glacial loess deposits in Patagonia). Northward extension of sea ice also leads to a steeper meridional temperature gradient and to more efficient poleward transport. Therefore, Vostok dust peaks would correspond to periods of increased sea-ice extent in the South Atlantic Ocean, probably associated with reduced deep ocean circulation (thus explaining observed similarities with the tropical ocean dust record³³).

Greenhouse gases. The extension of the greenhouse-gas record shows that the main trends of CO_2 and CH_4 concentration changes are similar for each glacial cycle (Fig. 3). Major transitions from the lowest to the highest values are associated with glacial–interglacial transitions. At these times, the atmospheric concentrations of CO_2 rises from 180 to 280–300 p.p.m.v. and that of CH_4 rises from 320–350 to 650–770 p.p.b.v. There are significant differences between the CH_4 concentration change associated with deglaciations. Termination III shows the smallest CH_4 increase, whereas termination IV shows the largest (Fig. 5). Differences in the changes over deglaciations are less significant for CO_2 . The decrease of CO_2 to the minimum values of glacial times is slower than its increase towards interglacial levels, confirming the sawtooth record of this property. CH_4 also decreases slowly to its background level, but with a series of superimposed peaks whose amplitude decreases during the course of each glaciation. Each CH_4 peak is itself characterized by rapid increases and slower decreases, but our resolution is currently inadequate to capture the detail of millennial-scale CH_4 variations. During glacial inception, Antarctic temperature and CH_4 concentrations decrease in phase. The CO_2 decrease lags the temperature decrease by several kyr and may be either steep (as at the end of interglacials 5.5 and 7.5) or more regular (at the end of interglacials 9.3 and 11.3). The differences in concentration–time profiles of CO_2 and CH_4 are reflected in the power spectra (Fig. 4). The 100-kyr component dominates both CO_2 and CH_4 records. However, the obliquity and precession components are much stronger for CH_4 than for CO_2 .

The extension of the greenhouse-gas record shows that present-day levels of CO_2 and CH_4 (~ 360 p.p.m.v. and $\sim 1,700$ p.p.b.v., respectively) are unprecedented during the past 420 kyr. Pre-industrial Holocene levels (~ 280 p.p.m.v. and ~ 650 p.p.b.v., respectively) are found during all interglacials, while values higher than these are found in stages 5.5, 9.3 and 11.3 (this last stage is probably incomplete), with the highest values during stage 9.3 (300 p.p.m.v. and 780 p.p.b.v., respectively).

The overall correlation between our CO_2 and CH_4 records and the Antarctic isotopic temperature^{5,9,16} is remarkable ($r^2 = 0.71$ and 0.73 for CO_2 and CH_4 , respectively). This high correlation indicates that CO_2 and CH_4 may have contributed to the glacial–interglacial changes over this entire period by amplifying the orbital forcing along with albedo, and possibly other changes^{5,16}. We have calculated the direct radiative forcing corresponding to the CO_2 , CH_4 and N_2O changes¹⁶. The largest CO_2 change, which occurs between stages 10 and 9, implies a direct radiative warming of

Table 1 Comparison of the glaciological timescale and orbitally derived information

Depth (m)	Insolation maximum (kyr)	Age GT4 (kyr)	Difference (kyr)
305	11	10	1
900	58	57	1
1,213	84	83	1
1,528	105	105	0
1,863	128	128	0
2,110	151	150	1
2,350	176	179	–3
2,530	199	203	–4
2,683	220	222	–2
2,788	244	239	5
2,863	265	255	10
2,972	293	282	11
3,042	314	301	13
3,119	335	322	13

Control points were derived assuming a correspondence between maximum 65°N mid-June insolation and $\delta^{18}\text{O}_{\text{atm}}$ mid-transitions. Age GT4 refers to the age of the gas obtained after correction for the gas-age/ice-age differences²⁰.

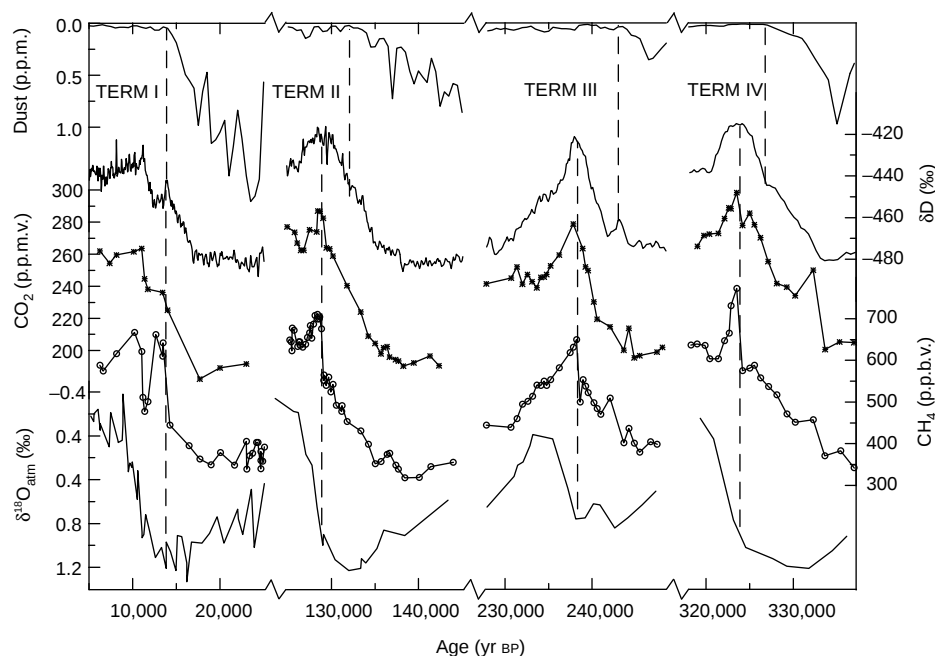


Figure 5 Vostok time series during glacial terminations. Variations with respect to time (GT4) of: **a**, dust; **b**, δD of ice (temperature proxy); **c**, CO_2 ; **d**, CH_4 ; and **e**, $\delta^{18}O_{atm}$

for glacial terminations I to IV and the subsequent interglacial periods (Holocene, stage 5.5, stage 7.5 and stage 9.3).

$\Delta T_{rad} = 0.75^\circ C$. Adding the effects of CH_4 and N_2O at this termination increases the forcing to $0.95^\circ C$ (here we assume that N_2O varies with climate as during termination I³⁷). This initial forcing is amplified by positive feedbacks associated with water vapour, sea ice, and possibly clouds (although in a different way for a 'doubled CO_2 ' situation than for a glacial climate³⁸). The total glacial–interglacial forcing is important ($\sim 3 W m^{-2}$), representing 80% of that corresponding to the difference between a 'doubled CO_2 ' world and modern CO_2 climate. Results from various climate simulations^{39,40} make it reasonable to assume that greenhouse gases have, at a global scale, contributed significantly (possibly about half, that is, $2\text{--}3^\circ C$) to the globally averaged glacial–interglacial temperature change.

Glacial terminations and interglacials

Our complete Vostok data set allows us to examine all glacial commencements and terminations of the past 420 kyr. We can examine the time course of the following five properties during these events: δD_{ice} , dust content, $\delta^{18}O_{atm}$, CO_2 and CH_4 (Fig. 5). We consider that, during the terminations, $\delta^{18}O_{atm}$ tracks $\delta^{18}O_{sw}$ with a lag of ~ 2 kyr (ref. 11), the response time of the atmosphere to changes in $\delta^{18}O_{sw}$. $\delta^{18}O_{atm}$ can thus be taken as an indicator of the large ice-volume changes associated with the deglaciations. Broecker and Henderson⁴¹ recently supported this interpretation for the last two terminations and discussed its limitations. Our extended $\delta^{18}O_{atm}$ record indeed reinforces such an interpretation, as it shows that the amplitudes of $\delta^{18}O_{atm}$ changes parallel $\delta^{18}O_{sw}$ changes for all four terminations. $\delta^{18}O_{sw}$ changes are similar for terminations I, II and IV ($1.1\text{--}1.2\text{‰}$) but much smaller for termination III ($\sim 0.6\text{‰}$). The same is true for $\delta^{18}O_{atm}$ ($1.4\text{--}1.5\text{‰}$ for I, II and IV, and 0.8‰ for III).

A striking feature of the Vostok deuterium record is that the Holocene, which has already lasted 11 kyr, is, by far, the longest stable warm period recorded in Antarctica during the past 420 kyr (Fig. 5). Interglacials 5.5 and 9.3 are different from the Holocene, but similar to each other in duration, shape and amplitude. During each of these two events, there is a warm period of ~ 4 kyr followed by a relatively rapid cooling and then a slower temperature decrease (Fig. 5), rather like some North Atlantic deep-sea core records⁴². Stage 7.5 is different in all respects, with a slightly colder maximum,

a more spiky shape, and a much shorter duration (7 kyr at mid-transition compared with 17 and 20 kyr for stages 5.5 and 9.3, respectively). This difference between stage 7.5 and stages 5.5 or 9.3 may result from the different configuration of the Earth's orbit (in particular concerning the phase of precession with respect to obliquity³). Termination III is also peculiar as far as terrestrial aerosol fallout is concerned. Terminations I, II and IV are marked by a large decrease in dust; high glacial values drop to low interglacial values by the mid-point of the δD_{ice} increase. But for termination III, the dust concentration decreases much earlier, with low interglacial values obtained just before a slight cooling event, as for termination I (for which interglacial values are reached just before the 'Antarctic Cold Reversal').

Unlike termination I, other terminations show, with our present resolution, no clear temperature anomalies equivalent to the Antarctic Cold Reversal⁶ (except possibly at the very beginning of termination III). There are also no older counterparts to the Younger Dryas CH_4 minimum⁴³ during terminations II, III and IV given the present resolution of the CH_4 record (which is no better than 1,000–2,000 yr before stage 5).

The sequence of events during terminations III and IV is the same as that previously observed for terminations I and II. Vostok temperature, CO_2 and CH_4 increase in phase during terminations. Uncertainty in the phasing comes mainly from the sampling frequency and the ubiquitous uncertainty in gas-age/ice-age differences (which are well over ± 1 kyr during glaciations and terminations). In a recent paper, Fischer *et al.*⁴⁴ present a CO_2 record, from Vostok core, spanning the past three glacial terminations. They conclude that CO_2 concentration increases lagged Antarctic warmings by 600 ± 400 years. However, considering the large gas-age/ice-age uncertainty (1,000 years, or even more if we consider the accumulation-rate uncertainty), we feel that it is premature to infer the sign of the phase relationship between CO_2 and temperature at the start of terminations. We also note that their discussion relates to early deglacial changes, not the entire transitions.

An intriguing aspect of the deglacial CH_4 curves is that the atmospheric concentration of CH_4 rises slowly, then jumps to a maximum value during the last half of the deglacial temperature rise. For termination I, the CH_4 jump corresponds to a rapid Northern Hemisphere warming (Bölling/Allerød) and an increase

in the rate of Northern Hemisphere deglaciation (meltwater pulse IA)⁴³. We speculate that the same is true for terminations II, III and IV. Supportive evidence comes from the $\delta^{18}\text{O}_{\text{atm}}$ curves. During each termination, $\delta^{18}\text{O}_{\text{atm}}$ begins falling rapidly, signalling intense deglaciation, within 1 kyr of the CH_4 jump. The lag of deglaciation and Northern Hemisphere warming with respect to Vostok temperature warming is apparently greater during terminations II and IV (~9 kyr) than during terminations I and III (~4–6 kyr). The changes in northern summer insolation maxima are higher during terminations II and IV, whereas the preceding southern summer insolation maxima are higher during terminations I and III. We speculate that variability in phasing from one termination to the next reflects differences in insolation curves⁴¹ or patterns of abyssal circulation during glacial maximum. Our results suggest that the same sequence of climate forcings occurred during each termination: orbital forcing (possibly through local insolation changes, but this is speculative as we have poor absolute dating) followed by two strong amplifiers, with greenhouse gases acting first, and then deglaciation enhancement via ice-albedo feedback. The end of the deglaciation is then characterized by a clear CO_2 maximum for terminations II, III and IV, while this feature is less marked for the Holocene.

Comparison of CO_2 atmospheric concentration changes with variations of other properties illuminates oceanic processes influencing glacial–interglacial CO_2 changes. As already noted for terminations I and II⁴¹, the sequence of climate events described above rules out the possibility that rising sea level induces the CO_2 increase at the beginning of terminations. On the other hand, the small CO_2 variations associated with Heinrich events⁴⁵ suggest that the formation of North Atlantic Deep Water does not have a large effect on CO_2 concentrations. Our record shows similar relative amplitudes of atmospheric CO_2 and Vostok temperature changes for the four terminations. Also, values of both CO_2 and temperature are significantly higher during stage 7.5 than during stages 7.1 and 7.3, whereas the deep-sea core ice volume record exhibits similar levels for these three stages. These similarities between changes in atmospheric CO_2 and Antarctic temperature suggest that the oceanic area around Antarctica plays a role in the long-term CO_2 change. An influence of high southern latitudes is also suggested by the comparison with the dust profile, which exhibits a maximum during the periods of lowest CO_2 . The link between dust and CO_2 variations could be through the atmospheric input of iron⁴⁶. Alternatively, we suggest a link through deep ocean circulation and sea ice extent in the Southern Ocean, both of which play a role in ocean CO_2 ventilation and, as suggested above, in the dust input over East Antarctica.

New constraints on past climate change

As judged from Vostok records, climate has almost always been in a state of change during the past 420 kyr but within stable bounds (that is, there are maximum and minimum values of climate properties between which climate oscillates). Significant features of the most recent glacial–interglacial cycle are observed in earlier cycles. Spectral analysis emphasises the dominance of the 100-kyr cycle for all six data series except $\delta^{18}\text{O}_{\text{atm}}$ and a strong imprint of 40- and/or 20-kyr periodicities despite the fact that the glaciological dating is tuned by fitting only two control points in the 100-kyr band.

Properties change in the following sequence during each of the last four glacial terminations, as recorded in Vostok. First, the temperature and atmospheric concentrations of CO_2 and CH_4 rise steadily, whereas the dust input decreases. During the last half of the temperature rise, there is a rapid increase in CH_4 . This event coincides with the start of the $\delta^{18}\text{O}_{\text{atm}}$ decrease. We believe that the rapid CH_4 rise also signifies warming in Greenland, and that the deglacial $\delta^{18}\text{O}_{\text{atm}}$ decrease records rapid melting of the Northern Hemisphere ice sheets. These results suggest that the same sequence

of climate forcing operated during each termination: orbital forcing (with a possible contribution of local insolation changes) followed by two strong amplifiers, greenhouse gases acting first, then deglaciation and ice-albedo feedback. Our data suggest a significant role of the Southern Ocean in regulating the long-term changes of atmospheric CO_2 .

The Antarctic temperature was warmer, and atmospheric CO_2 and CH_4 concentrations were higher, during interglacials 5.5 and 9.3 than during the Holocene and interglacial 7.5. The temporal evolution and duration of stages 5.5 and 9.3 are indeed remarkably similar for all properties recorded in Vostok ice and entrapped gases. As judged from the Vostok record, the long, stable Holocene is a unique feature of climate during the past 420 kyr, with possibly profound implications for evolution and the development of civilizations. Finally, CO_2 and CH_4 concentrations are strongly correlated with Antarctic temperatures; this is because, overall, our results support the idea that greenhouse gases have contributed significantly to the glacial–interglacial change. This correlation, together with the uniquely elevated concentrations of these gases today, is of relevance with respect to the continuing debate on the future of Earth's climate.

Received 20 January; accepted 14 April 1999.

1. Imbrie, J. *et al.* On the structure and origin of major glaciation cycles. 1. Linear responses to Milankovich forcing. *Paleoceanography* **7**, 701–738 (1992).
2. Tzedakis, P. C. *et al.* Comparison of terrestrial and marine records of changing climate of the last 500,000 years. *Earth Planet. Sci. Lett.* **150**, 171–176 (1997).
3. Berger, A. L. Long-term variations of daily insolation and Quaternary climatic change. *J. Atmos. Sci.* **35**, 2362–2367 (1978).
4. Lorius, C. *et al.* A 150,000-year climatic record from Antarctic ice. *Nature* **316**, 591–596 (1985).
5. Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. Vostok ice cores provides 160,000-year record of atmospheric CO_2 . *Nature* **329**, 408–414 (1987).
6. Jouzel, J. *et al.* Vostok ice core: a continuous isotope temperature record over the last climatic cycle (160,000 years). *Nature* **329**, 402–408 (1987).
7. Raisbeck, G. M. *et al.* Evidence for two intervals of enhanced ^{10}Be deposition in Antarctic ice during the last glacial period. *Nature* **326**, 273–277 (1987).
8. Legrand, M., Lorius, C., Barkov, N. I. & Petrov, V. N. Vostok (Antarctic ice core): atmospheric chemistry changes over the last climatic cycle (160,000 years). *Atmos. Environ.* **22**, 317–331 (1988).
9. Chappellaz, J., Barnola, J.-M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. Ice-core record of atmospheric methane over the past 160,000 years. *Nature* **329**, 127–131 (1990).
10. Petit, J. R. *et al.* Paleoclimatological implications of the Vostok core dust record. *Nature* **343**, 56–58 (1990).
11. Sowers, T. *et al.* 135 000 year Vostok—SPECMAP common temporal framework. *Paleoceanography* **8**, 737–766 (1993).
12. Jouzel, J. *et al.* Extending the Vostok ice-core record of palaeoclimate to the penultimate glacial period. *Nature* **364**, 407–412 (1993).
13. Jouzel, J. *et al.* Climatic interpretation of the recently extended Vostok ice records. *Clim. Dyn.* **12**, 513–521 (1996).
14. Genthon, C. *et al.* Vostok ice core: climatic response to CO_2 and orbital forcing changes over the last climatic cycle. *Nature* **329**, 414–418 (1987).
15. Lorius, C., Jouzel, J., Raynaud, D., Hansen, J. & Le Treut, H. Greenhouse warming, climate sensitivity and ice core data. *Nature* **347**, 139–145 (1990).
16. Raynaud, D. *et al.* The ice record of greenhouse gases. *Science* **259**, 926–934 (1993).
17. Petit, J. R. *et al.* Four climatic cycles in Vostok ice core. *Nature* **387**, 359–360 (1997).
18. Kapitzka, A. P., Ridley, J. K., Robin, G. d. Q., Siegert, M. J. & Zotikov, I. A. A large deep freshwater lake beneath the ice of central East Antarctica. *Nature* **381**, 684–686 (1996).
19. Bender, M., Sowers, T. & Labeyrie, L. D. The Dole effect and its variation during the last 130,000 years as measured in the Vostok core. *Glob. Biogeochem. Cycles* **8**, 363–376 (1994).
20. Barnola, J. M., Pimienta, P., Raynaud, D. & Korotkevich, Y. S. CO_2 climate relationship as deduced from the Vostok ice core: a re-examination based on new measurements and on a re-evaluation of the air dating. *Tellus B* **43**, 83–91 (1991).
21. Lorius, C. & Merlivat, L. in *Isotopes and Impurities in Snow and Ice. Proc. the Grenoble Symp. Aug./Sept. 1975* 127–137 (Publ. 118, IAHS, 1977).
22. Dahl-Jensen, D. *et al.* Past temperatures directly from the Greenland ice sheet. *Science* **282**, 268–271 (1998).
23. Salamin, A. N. *et al.* Ice core age dating and paleothermometer calibration on the basis of isotopes and temperature profiles from deep boreholes at Vostok station (East Antarctica). *J. Geophys. Res.* **103**, 8963–8977 (1998).
24. Krinner, G., Genthon, C. & Jouzel, J. GCM analysis of local influences on ice core δ signals. *Geophys. Res. Lett.* **24**, 2825–2828 (1997).
25. Hoffmann, G., Masson, V. & Jouzel, J. Stable water isotopes in atmospheric general circulation models. *Hydrol. Processes* (in the press).
26. Bassinot, F. C. *et al.* The astronomical theory of climate and the age of the Brunhes-Matuyama magnetic reversal. *Earth Planet. Sci. Lett.* **126**, 91–108 (1994).
27. Shackleton, N. J., Imbrie, J. & Hall, M. A. Oxygen and carbon isotope record of East Pacific core V19-30: implications for the formation of deep water in the late Pleistocene North Atlantic. *Earth Planet. Sci. Lett.* **65**, 233–244 (1983).
28. Steig, E. *et al.* Synchronous climate changes in Antarctica and the North Atlantic. *Science* **282**, 92–95 (1998).
29. Howard, W. A warm future in the past. *Nature* **388**, 418–419 (1997).
30. Malaizé, B., Paillard, D., Jouzel, J. & Raynaud, D. The Dole effect over the last two glacial-interglacial cycles. *J. Geophys. Res.* (in the press).
31. Legrand, M. & Delmas, R. J. Formation of HCl in the Antarctic atmosphere. *J. Geophys. Res.* **93**, 7153–7168 (1987).
32. Yung, Y. K., Lee, T., Chung-Ho & Shieh, Y. T. Dust: diagnostic of the hydrological cycle during the last glacial maximum. *Science* **271**, 962–963 (1996).

33. de Menocal, P. Plio-Pleistocene African climate. *Science* **270**, 53–59 (1995).
34. CLIMAP. *Seasonal Reconstructions of the Earth's Surface at the Last Glacial Maximum* (Geol. Soc. Am., Boulder, Colorado, 1981).
35. Basile, I. *et al.* Patagonian origin dust deposited in East Antarctica (Vostok and Dome C) during glacial stages 2, 4 and 6. *Earth Planet. Sci. Lett.* **146**, 573–589 (1997).
36. Basile, I. *Origine des Aérosols Volcaniques et Continentaux de la Carotte de Glace de Vostok (Antarctique)*. Thesis, Univ. Joseph Fourier, Grenoble (1997).
37. Leuenberger, M. & Siegenthaler, U. Ice-age atmospheric concentration of nitrous oxide from an Antarctic ice core. *Nature* **360**, 449–451 (1992).
38. Ramstein, G., Serafini-Le Treut, Y., Le Treut, H., Forichon, M. & Joussaume, S. Cloud processes associated with past and future climate changes. *Clim. Dyn.* **14**, 233–247 (1998).
39. Berger, A., Loutre, M. F. & Gallée, H. Sensitivity of the LLN climate model to the astronomical and CO₂ forcings over the last 200 ky. *Clim. Dyn.* **14**, 615–629 (1998).
40. Weaver, A. J., Eby, M., Fanning, A. F. & Wilbe, E. C. Simulated influence of carbon dioxide, orbital forcing and ice sheets on the climate of the Last Glacial Maximum. *Nature* **394**, 847–853 (1998).
41. Broecker, W. S. & Henderson, G. M. The sequence of events surrounding termination II and their implications for the causes of glacial interglacial CO₂ changes. *Paleoceanography* **13**, 352–364 (1998).
42. Cortijo, E. *et al.* Eemian cooling in the Norwegian Sea and North Atlantic ocean preceding ice-sheet growth. *Nature* **372**, 446–449 (1994).
43. Chappellaz, J. *et al.* Synchronous changes in atmospheric CH₄ and Greenland climate between 40 and 8 kyr BP. *Nature* **366**, 443–445 (1993).
44. Fischer, H., Wahlen, M., Smith, J., Mastoianni, D. & Deck, B. Ice core records of atmospheric CO₂ around the last three glacial terminations. *Science* **283**, 1712–1714 (1999).
45. Stauffer, B. *et al.* Atmospheric CO₂ concentration and millennial-scale climate change during the last glacial period. *Nature* **392**, 59–61 (1998).
46. Martin, J. H. Glacial-interglacial CO₂ change: The iron hypothesis. *Paleoceanography* **5**, 1–13 (1990).
47. Paillard, D., Labeyrie, L. & Yiou, P. Macintosh program performs time-series analysis. *Eos* **77**, 379 (1996).
48. Ritz, C. *Un Modèle Thermo-mécanique d'évolution Pour le Bassin Glaciaire Antarctique Vostok-Glacier Byrd: Sensibilité aux Valeurs de Paramètres Mal Connus* Thesis, Univ. Grenoble (1992).
49. Blunier, T. *et al.* Timing of the Antarctic Cold Reversal and the atmospheric CO₂ increase with respect to the Younger Dryas event. *Geophys. Res. Lett.* **24**, 2683–2686 (1997).
50. Waelbroeck, C. *et al.* A comparison of the Vostok ice deuterium record and series from Southern Ocean core MD 88-770 over the last two glacial-interglacial cycles. *Clim. Dyn.* **12**, 113–123 (1995).
51. Blunier, T. *et al.* Asynchrony of Antarctic and Greenland climate change during the last glacial period. *Nature* **394**, 739–743 (1998).
52. Bender, M., Malaizé, B., Orchado, J., Sowers, T. & Jouzel, J. High precision correlations of Greenland and Antarctic ice core records over the last 100 kyr. in *The Role of High and Low Latitudes in Millennial Scale Global Change* (eds Clark, P. & Webb, R.) (AGU Monogr., Am. Geophys. Union, in the press).

Supplementary information is available on *Nature's* World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of *Nature*.

Acknowledgements. This work is part of a joint project between Russia, France and USA. We thank the drillers from the St Petersburg Mining Institute; the Russian, French and US participants for field work and ice sampling; and the Russian Antarctic Expeditions (RAE), the Institut Français de Recherches et Technologies Polaires (IFRTP) and the Division of Polar Programs (NSF) for the logistic support. The project is supported in Russia by the Russian Ministry of Sciences, in France by PNEDC (Programme National d'Études de la Dynamique du Climat), by Fondation de France and by the CEC (Commission of European Communities) Environment Programme, and in the US by the NSF Science Foundation.

Correspondence and requests for materials should be addressed to J.R.P. (e-mail: petit@glaciog.ujf-grenoble.fr).

A doubling of the Sun's coronal magnetic field during the past 100 years

M. Lockwood, R. Stamper & M. N. Wild

World Data Centre C-1 for STP, Rutherford Appleton Laboratory, Chilton, Didcot OX11 0QX, UK

The solar wind is an extended ionized gas of very high electrical conductivity, and therefore drags some magnetic flux out of the Sun to fill the heliosphere with a weak interplanetary magnetic field^{1,2}. Magnetic reconnection—the merging of oppositely directed magnetic fields—between the interplanetary field and the Earth's magnetic field allows energy from the solar wind to enter the near-Earth environment. The Sun's properties, such as its luminosity, are related to its magnetic field, although the connections are still not well understood^{3,4}. Moreover, changes in the

heliospheric magnetic field have been linked with changes in total cloud cover over the Earth, which may influence global climate⁵. Here we show that measurements of the near-Earth interplanetary magnetic field reveal that the total magnetic flux leaving the Sun has risen by a factor of 1.4 since 1964: surrogate measurements of the interplanetary magnetic field indicate that the increase since 1901 has been by a factor of 2.3. This increase may be related to chaotic changes in the dynamo that generates the solar magnetic field. We do not yet know quantitatively how such changes will influence the global environment.

The 'aa' index has been compiled from the range of variations in the geomagnetic field over periods of 3 hours, recorded since 1868 by pairs of near-antipodal magnetometers in England and Australia (see ref. 6 for a complete description of the index). Figure 1 demonstrates that the annual means (aa) show a marked variation with the sunspot cycle, but have also drifted upward throughout

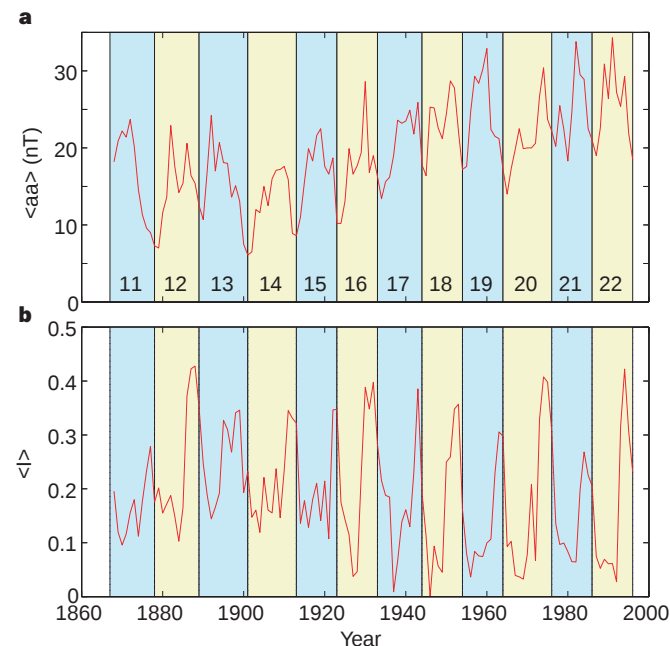


Figure 1 Annual means of **a**, the geomagnetic activity index, (aa), and **b**, Sargent's recurrence index, (I). I is defined for the jth 27-day Carrington rotation period as $I_j = (1/13) \sum_{k=-6}^{+6} C(j+k, j+k+1)$, where c is the correlation coefficient between two consecutive 27-day intervals of 12-hourly aa values^{13,23}. The data are for 1868–1996, covering sunspot cycles 11–22: alternate solar cycles are shaded yellow and blue, starting at sunspot minima. Non-recurrent increases in aa are caused by solar disturbances, such as coronal mass ejections, hitting the Earth's magnetosphere. This occurs more frequently at sunspot maximum²⁵. In the declining phase of each sunspot cycle, (aa) is high because Earth intersects long-lived, fast solar-wind streams¹³. These emanate from coronal holes that have expanded to low heliospheric latitudes²² and rotate with the equatorial photosphere every 25 days. During this time, the Earth moves along its orbit, such that it intersects the same stream every 27 days giving recurrent geomagnetic storms and high (I) (ref. 13). Several long-term trends are apparent: for all phases of the solar cycle, (aa) increased gradually between 1900 and about 1955, before decreasing and then rising again during 1964–96. Similar trends can be seen in the sunspot number maxima (see Fig. 3). The recurrent element of geomagnetic activity at sunspot maximum, as quantified by (I), fell during the early part of the century, but the sunspot minimum/declining phase peaks in (I) remained high with a tendency for greater values during even-numbered sunspot cycles²⁴.

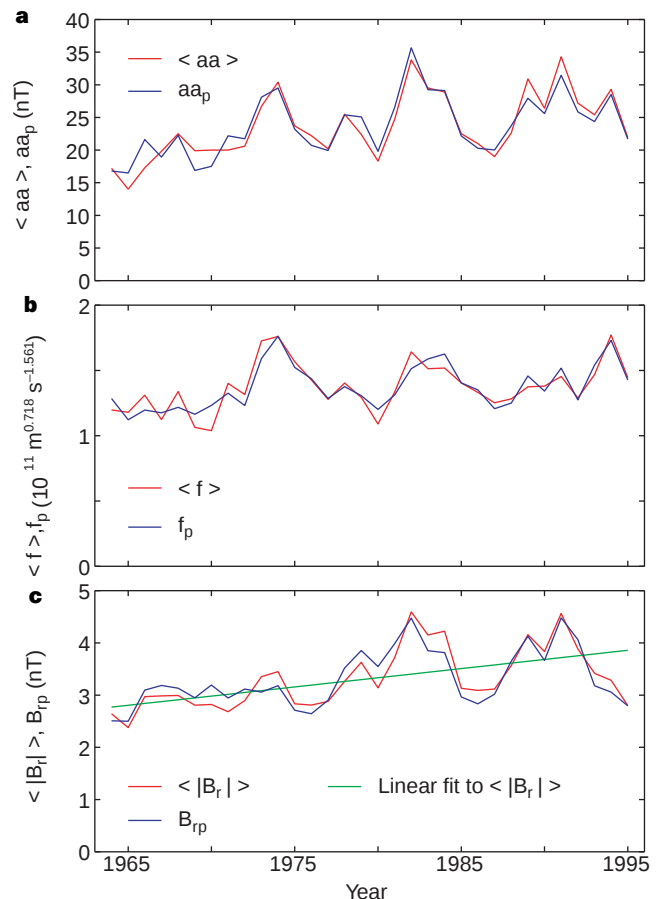


Figure 2 Time series of observed annual means and corresponding best-fit predicted values for 1964–96. **a**, Observed (aa) and predicted $aa_p = s_a P_a$ for the optimum coupling exponent α of 0.386 (ref. 7): a least-squares linear regression fit yields s_a , such that aa_p (in nT) = $(5.317 \times 10^{-17}) \{M_E \text{ in } T m^3\}^{2/3} \{N_{sw} \text{ in } m^{-3}\}^{0.281} \{v_{sw} \text{ in } km s^{-1}\}^{0.561} \{B_{sw} \text{ in } nT\}^{0.772} \langle \sin^4(\theta/2) \rangle$. The correlation coefficient is 0.94, giving a significance level of $(100 - (1.3 \times 10^{-13}))\%$. There are larger uncertainties in the calibration of the interplanetary data, particularly N_{sw} before 1974¹: excluding these raises the correlation coefficient to 0.97, but lowers the significance level slightly to $(100 - (1.3 \times 10^{-11}))\%$. **b**, The annual means $\langle I \rangle = \langle v_{sw} \text{ in } m^{-3} \rangle^{0.281} \{v_{sw} \text{ in } km s^{-1}\}^{0.561} \langle \sin^4(\theta/2) \rangle$ and the best-fit predicted value, $I_p = s_i \langle I \rangle^\beta (aa)^\lambda + c_i$. The best-fit constants are $\beta = 0.263$, $\lambda = 1.303$, $s_i = 2.607 \times 10^4$ and $c_i = 1.893 \times 10^6$. The correlation coefficient is 0.91, for which the significance level is $(100 - (4.3 \times 10^{-11}))\%$. **c**, The annual means of the amplitude of the radial IMF component (B_r) and the predicted value $B_{rp} = s_b \langle B_{sw} \rangle$, where B_{sw} is the IMF magnitude. The least-squares fit gives the slope $s_b = 0.56$. The correlation coefficient is 0.92, for which the significance level is $(100 - (2.5 \times 10^{-12}))\%$. The green line in **c** is a linear regression fit to $\langle B_r \rangle$.

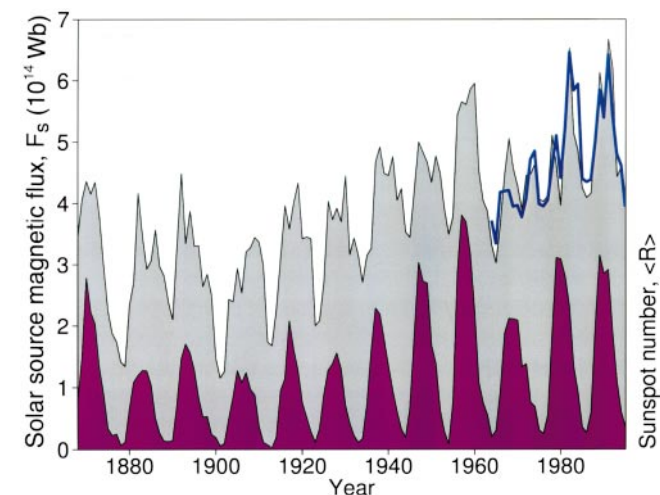


Figure 3 The total solar magnetic flux emanating through the coronal source sphere¹², F_s . Shown are the values derived from the geomagnetic aa data for 1868–1996 (black line bounding grey shading) and the values from the interplanetary observations for 1964–96 (thick blue line). The variation of the annual means of the sunspot number $\langle R \rangle$ is shown by the area shaded purple and varies between 0 and a peak of 190 for solar cycle 19.

most of this century. These changes are almost entirely due to variations in near-Earth interplanetary space⁷. Several attempts have been made to use the aa data to deduce the interplanetary and solar conditions before the ‘space age’^{8–10}. The success of these extrapolations depends critically on the quality of the correlation found between aa and the combination of the interplanetary parameters (the empirical ‘coupling function’¹¹) used to quantify the controlling influence of the solar wind and the interplanetary magnetic field (IMF). Recently, an unprecedentedly high and significant correlation coefficient of 0.97 has been obtained⁷.

We derive information on the magnetic field in the solar atmosphere (corona) from the aa data, using a procedure described in Methods. F_s is the magnetic flux that threads a roughly spherical ‘source’ surface in the corona where the Sun’s field becomes purely radial: it quantifies the amount of flux leaving the Sun and entering the heliosphere. The method employs three correlations of extremely high significance for the period 1964–96. Figure 2 shows the good agreement between observed yearly averages and their best-fit predicted values, derived using these correlations. Figure 2a shows the observed and predicted annual means of aa ($\langle aa \rangle$ and aa_p , respectively), Fig. 2b is the same for the function f (see Methods), and Fig. 2c is for the magnitude of the radial component of the IMF $|B_r|$.

The orientation of the IMF in annual averages is as predicted by the theory^{1,7} of the ‘Parker spiral’. This theory also predicts that the rise in the magnitude of the mean radial component (Fig. 2c) reflects a corresponding change in the coronal source field B_0 (equation (5)). A least-squares linear fit to $|B_r|$ for 1964–96 (the green line in Fig. 2c) yields a percentage change (defined as 100 times the change, divided by the initial value) of 41% ($\pm 13\%$). In other words, there has been a rise by a factor of 1.41 over the last three solar cycles. This rise is present, but not commented on, in previously published coronal source field estimates, modelled from the measured solar photospheric field¹². Cosmic rays are shielded from the Earth by both the IMF and the solar-wind flow, and the observed decay in cosmic-ray fluxes (by 3.7% since 1964)^{7,13} is, at least qualitatively, consistent with the rise in the IMF.

The results of the extrapolation to before 1964 are shown in Fig. 3. The values of F_s derived from the aa data are shown in grey, and compare well with those from the observed annual means of the IMF radial component $|B_r|$ for 1964–96 (thick blue line). The coronal source flux rises and falls in each solar cycle, lagging only

slightly behind the sunspot numbers R , shown in purple. The main differences between F_s and $\langle aa \rangle$ arise because the effects of the recurrent fast solar-wind streams (in the declining phase of each cycle¹³) have effectively been removed by our procedure. For data at all phases of the solar cycle, F_s has a correlation coefficient of 0.75 with the simultaneous R (giving a significance level of effectively 100%). To eliminate the solar-cycle variations, we have studied the 11-year running means and those for F_s and R vary in a very similar way. In 1901, the 11-year running mean of F_s was a minimum of 2.308×10^{14} Wb, but rose to a peak value of 5.325×10^{14} Wb in 1992. Thus in the intervening 91 years (covering roughly 8.5 sunspot cycles) there was a rise in the average solar source flux of 131% (that is, a rise by a factor of 2.31).

These changes in the solar magnetic field should be seen in the context of longer-term changes in the Sun, as inferred from historical sunspot and auroral observations¹⁴ and from the terrestrial abundances of isotopes such as ^{14}C and ^{10}Be (produced by cosmic-ray bombardment and deposited and stored, for example, in the polar icecaps)^{15,16}. The isotope data show that solar activity can largely disappear for periods of 50–100 years; one such period was the ‘Maunder minimum’ (circa 1645–1715), although there is evidence that during this interval a weak and cyclic magnetic field still emerged from the Sun¹⁶. By comparing the phase of the 88-year oscillation before and after the Maunder minimum, it has been inferred that the dynamo generating the solar field may be chaotic rather than quasi-periodic¹⁷: such behaviour may be relevant to the sudden changes in F_s around 1900 and 1960. Recent studies have linked changes in solar activity and aa with terrestrial climate change^{3–5,18}. The variation found here stresses the importance of understanding the connections between the Sun’s output and its magnetic field^{3,4} and between terrestrial global cloud cover, cosmic-ray fluxes and the heliospheric field⁵. □

Methods

We employ the optimum energy coupling function between the solar wind and the Earth’s magnetosphere derived by Stamper *et al.*⁷ using the dimensional analysis proposed by Vasyliunas *et al.*¹⁹. The solar-wind kinetic energy density dominates over the energy densities of both thermal motions and the IMF. This is incident on the geomagnetic field, which presents a roughly circular cross-section to the flow. A fraction of the incident energy is extracted, the power

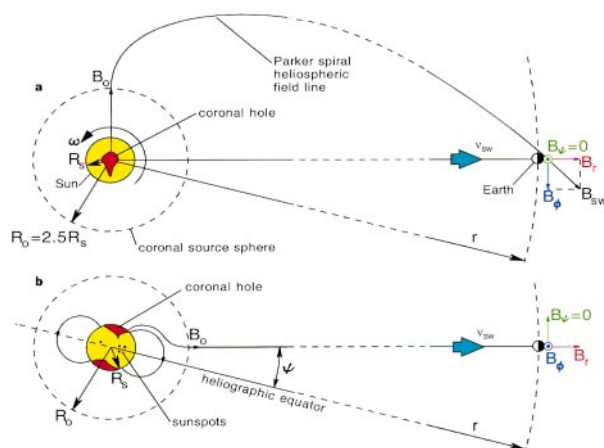


Figure 4 Schematic illustration of how the IMF B_{sw} emerges from holes in the solar atmosphere (coronal holes) and is dragged to Earth by the solar wind. The solar wind is flowing at v_{sw} radially away from the Sun in the heliosphere. The solar rotation winds the IMF into the Parker spiral¹. Such a field line is viewed here **a**, from north of the ecliptic plane and **b**, from a point in the ecliptic plane, to the dusk side of Earth. The coronal source surface is where the magnetic field, B_0 , is purely radial and is at a heliocentric distance of R_0 , which is roughly 2.5 times the solar radius, R_s (ref. 12). The magnetic flux threading this surface is F_s . The Ulysses spacecraft has shown that the radial component of the heliospheric field, B_r , is almost independent of the heliographic latitude ψ , (ref. 2).

transferred to the magnetosphere being⁷:

$$P_{\alpha} = \{k\pi/2\mu_0^{(1/3-\alpha)}\}M_E^{2/3}m_{sw}^{(2/3-\alpha)}N_{sw}^{(2/3-\alpha)}v_{sw}^{(7/3-2\alpha)}B_{sw}^{2\alpha}\sin^4(\theta/2) = aa_p/s_a \quad (1)$$

where m_{sw} is the mean ion mass, N_{sw} the concentration and v_{sw} the speed of the solar wind. B_{sw} is the IMF magnitude, θ is the IMF orientation "clock angle"²⁰, M_E is the magnetic moment of the Earth (taken from the IGRF model²¹), s_a and k are constants and aa_p is the best-fit prediction of aa . From annual means for 1964–96, the best-fit "coupling exponent" α is found to be 0.386 (ref. 7), and s_a is obtained from a linear regression fit of $\langle aa \rangle$ against P_{α} . The largest factor contributing to the rise in $\langle aa \rangle$ since 1964 is an upward drift in B_{sw} with significant rises in N_{sw} and v_{sw} ; however the mean θ has grown somewhat less favourable for increasing $\langle aa \rangle$ (ref. 7). The dependence is sufficient to allow derivation of B_{sw} from $\langle aa \rangle$. In order to separate the effect of B_{sw} from that of the other interplanetary variables, we define a parameter f :

$$f = N_{sw}^{(2/3-\alpha)}v_{sw}^{(7/3-2\alpha)}\sin^4(\theta/2) \quad (2)$$

the variation of which is dominated by that in the solar wind speed v_{sw} . The annual mean of v_{sw} rises in the declining phase of solar cycles¹³ because the Earth repeatedly intersects fast solar-wind streams from low-latitude extensions of coronal holes²². These occur every 27 days and so also raise the geomagnetic recurrence index, I (see Fig. 1)²³. Hence we expect f and I to increase together in the declining phase of the sunspot cycle. However, I can remain high at sunspot minimum (whereas v_{sw} is lower) because aa values are low and relatively constant²⁴. Hence we adopted a relationship for a predicted f of the form:

$$f_p = s_f I^{\beta} aa^{\lambda} + c_f \quad (3)$$

where the exponents β and λ give the optimum correlation, and the constants s_f and c_f are then found from a linear regression fit. The primary justification for the use of equation (3) is that it yields a correlation which is comparable (in magnitude and significance) to the other two shown in Fig. 2. Note that f_p reproduces both the drift and 22-year cycle in f . From equations (1)–(3) we can obtain a formula for estimating B_{sw} from the aa index data series:

$$B_{sw} = \{[2aa\mu_0^{(1/3-\alpha)}]/\{s_a k \pi m_{sw}^{(2/3-\alpha)} M_E^{2/3} (s_f I^{\beta} aa^{\lambda} + c_f)\}\}^{1/(2\alpha)} \quad (4)$$

Parker spiral theory successfully predicts the radial and latitudinal variations of the annual means of the heliospheric field¹⁷:

$$B_{sw} = \{B_r^2 + B_{\phi}^2 + B_{\psi}^2\}^{1/2} = B_r \{1 + \tan^2 \gamma\}^{1/2} = B_0 (R_0/r)^2 \{1 + (\omega r \cos \psi / v_{sw})^2\}^{1/2} \quad (5)$$

where B_0 is the coronal source field at R_0 from the centre of the Sun¹², ω is the equatorial angular solar rotation velocity and ψ is the heliographic latitude (Fig. 4). In annual means, the modulus of the out-of-ecliptic IMF component $\langle B_{\psi} \rangle$ is well correlated with B_{sw} and the mean $\langle B_{\psi} \rangle$ is close to zero. The 'garden hose angle' γ of the IMF in the ecliptic plane (equal to $\tan^{-1} B_{\phi}/B_r$) remains close to 45°, and so the radial heliospheric field component B_r is roughly proportional to B_{sw} , that is, $B_r = s_B B_{sw}$ (Fig. 2c). In addition, recent observations by the Ulysses satellite have shown that latitudinal variations in the heliospheric field are small (B_r is independent of ψ)². This result has been used to derive the coronal source field B_0 from photospheric field measurements, and good agreement found with observations of B_r near the Earth at all phases of the solar cycle¹². The total magnetic flux of the sun that threads the source surface (radius R_0), F_s is:

$$F_s = (1/2)4\pi R_0^2 B_0 = 2\pi r^2 B_r = 2\pi r^2 s_B B_{sw} \quad (6)$$

where $r = 1$ AU for observations near Earth¹². The factor of one-half arises because half the field threading the source surface is inward, the other half outward. We can compute the solar flux F_s from the aa data using equations (4) and (6).

To extrapolate to before 1964, we assume that all three correlations (derived from the data for after 1964) were valid at all times since 1868. Specifically, we assume that the empirical functions $\sin^4(\theta/2)$ and f_p behave as they did after 1964. We also assume that the empirical coupling exponent α and the mean mass of the solar wind are constant, that Parker spiral theory applied, and that latitudinal variations in the heliospheric field were small then as now.

Received 21 December 1998; accepted 12 April 1999.

1. Gazis, P. R. Solar cycle variation of the heliosphere. *Rev. Geophys.* **34**, 379–402 (1996).
2. Balogh, A. et al. The heliospheric field over the south polar region of the sun. *Science* **268**, 1007–1010 (1995).

3. Willson, R. C. Total solar irradiance trend during cycles 21 and 22. *Science* **277**, 1963–1965 (1997).
4. Lean, J., Beer, J. & Bradley, R. Reconstruction of solar irradiance since 1610: implications for climate change. *Geophys. Res. Lett.* **22**, 3195–3198 (1995).
5. Svensmark, H. & Friis-Christensen, E. Variation of cosmic ray flux and global cloud coverage—a missing link in solar-climate relationships. *J. Atmos. Sol. Terr. Phys.* **59**, 1225–1232 (1997).
6. Mayaud, P. N. The aa indices: a 100-year series characterising the magnetic activity. *J. Geophys. Res.* **72**, 6870–6874 (1972).
7. Stamper, R., Lockwood, M., Wild, M. N. & Clark, T. D. G. Solar causes of the long-term increase in geomagnetic activity. *J. Geophys. Res.* (in press).
8. Russell, C. T. On the possibility of deducing interplanetary and solar parameters from geomagnetic records. *Sol. Phys.* **42**, 259–269 (1975).
9. Gringauz, K. I. in *Solar Wind 4* (ed. Rosenbauer, H.) (Rep. MPAE-W-100-81-31, MPI für Aeronomie, Lindau, Germany, 1981).
10. Feynman, J. & Crooker, N. U. The solar wind at the turn of the century. *Nature* **275**, 626–627 (1978).
11. Baker, D. in *Solar Wind-Magnetosphere Coupling* (eds Kamide, Y. & Slavin, J. A.) 17–38 (Terra Scientific, Tokyo, 1986).
12. Wang, Y.-M. & Sheeley, N. R. Jr Solar implications of Ulysses interplanetary field measurements. *Astrophys. J.* **447**, L143–L146 (1995).
13. Cliver, E. W., Boriakoff, V. & Bounar, K. H. The 22-year cycle of geomagnetic activity. *J. Geophys. Res.* **101**, 27091–27109 (1996).
14. Silverman, S. W. Secular variation of the aurora for the past 500 years. *Rev. Geophys.* **30**, 333–351 (1992).
15. Sonnet, C. P. Long-period solar terrestrial variability. *Rev. Geophys. (Suppl.: US Nat. Rep. to IUGG 1987–1990)* 909–914 (1991).
16. Beer, J., Tobias, S. & Weiss, N. An active sun throughout the Maunder minimum. *Sol. Phys.* **181**, 237–249 (1998).
17. Feynman, J. & Gabriel, S. B. Period and phase of the 88-year solar cycle and the Maunder minimum: evidence for the chaotic sun. *Sol. Phys.* **127**, 393–403 (1990).
18. Cliver, E. W., Boriakoff, V. & Feynman, J. Solar variability and climate change: geomagnetic aa index and global surface temperature. *Geophys. Res. Lett.* **25**, 1035–1038 (1998).
19. Vasyliunas, V. M., Kan, J. R., Siscoe, G. L. & Akasofu, S.-I. Scaling relations governing magnetospheric energy transfer. *Planet Space Sci.* **30**, 359–365 (1982).
20. Scurry, L. & Russell, C. T. Proxy studies of energy transfer to the magnetosphere. *J. Geophys. Res.* **96**, 9541–9548 (1991).
21. Merrill, R. T., McElhinny, M. W. & McFadden, P. L. *The Magnetic Field of the Earth* 34 (Academic, San Diego, 1996).
22. Wang, Y.-M., Hawley, S. H. & Sheeley, N. R. Jr The magnetic nature of coronal holes. *Science* **271**, 464–469 (1996).
23. Sargent, H. H. in *Solar Wind-Magnetosphere Coupling* (eds Kamide, Y. & Slavin, J. A.) 143–148 (Terra Scientific, Tokyo, 1986).
24. Hapgood, M. A. A double solar-cycle variation in the 27-day recurrence of geomagnetic activity. *Ann Geophys.* **11**, 248–253 (1993).
25. Webb, D. F. & Howard, R. A. The solar cycle variation of coronal mass ejections and solar wind mass flux. *J. Geophys. Res.* **99**, 4201–4220 (1994).

Acknowledgements. The data used are stored and made available via World Data Centre C1 for STP at RAL, which is funded by the UK Particle Physics and Astronomy Research Council and, until 1 April 1999, by the National Radio Propagation Programme of the Radiocommunications Agency. We also thank the many scientists who have contributed data to the WDC.

Correspondence and requests for materials should be addressed to M.L. (e-mail: m.lockwood@rl.ac.uk).

Origin of high critical currents in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ superconducting thin films

B. Dam*, J. M. Huijbregtse*, F. C. Klaassen*, R. C. F. van der Geest*, G. Doornbos*, J. H. Rector*, A. M. Testa†, S. Freisem§, J. C. Martinez||, B. Stäuble-Pümpin¶ & R. Griessen*

* Institute COMPAS and Faculty of Sciences, Division of Physics and Astronomy, Vrije Universiteit, De Boelelaan 1081, NL-1081 HV Amsterdam, The Netherlands

† ICMAT, CNR, Area della Ricerca di Montelibretti, Rome, Italy

§ Kamerlingh Onnes Laboratory, Leiden University, PO Box 9506,

NL-2300 RA Leiden, The Netherlands

|| Institut für Physik, Johannes Gutenberg Universität, D-55099 Mainz, Germany

¶ Department of Physics, Faculty of Science, Universidad Nacional de Colombia,

Bogotá, Colombia

Thin films of the high-temperature superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ exhibit both a large critical current (the superconducting current density generally lies between 10^{11} and 10^{12} A m^{-2} at 4.2 K in zero magnetic field) and a decrease in such currents with magnetic field that point to the importance of strong vortex pinning along extended defects^{1,2}. But it has hitherto been unclear which types of defect—dislocations, grain boundaries, surface corrugations and anti-phase boundaries—are responsible. Here we make use of a sequential etching technique to address this question. We find that both edge and screw dislocations, which can be mapped

quantitatively by this technique, are the linear defects that provide the strong pinning centres responsible for the high critical currents observed in these thin films. Moreover, we find that the superconducting current density is essentially independent of the density of linear defects at low magnetic fields. These natural linear defects, in contrast to artificially generated columnar defects, exhibit self-organized short-range order, suggesting that $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ thin films offer an attractive system for investigating the properties of vortex matter in a superconductor with a tailored defect structure.

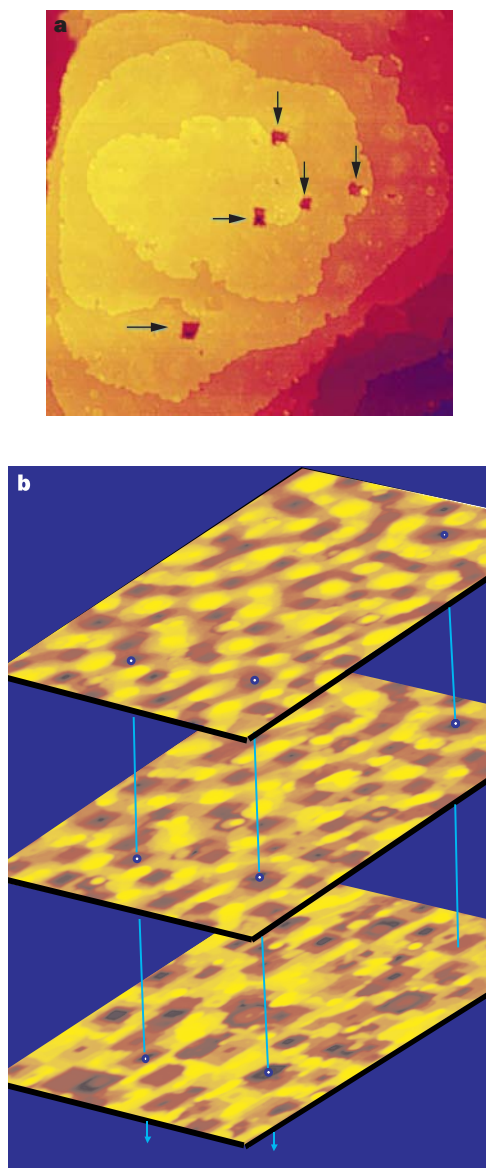


Figure 1 Surface morphology and etch-pitch distribution for sputtered and laser-deposited films, respectively. (These are AFM height images; bright is high, dark is low.) **a**, Large double spiral on a sputtered $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ film. After etching for 20 s, the growth features are still intact but small etch pits have developed at the spiral cores (screw dislocations, horizontal arrows) and randomly across the spiral (edge dislocations, vertical arrows). Image size $1 \times 1 \mu\text{m}$. **b**, Typical sequence of three AFM height images from a series of 25 frames, all taken from the same area of a pulsed laser deposited $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ film on repeated etching. After 15 s (top) the as-grown surface is still visible, but etch pits develop at the trenches between the growth islands; after 25 s (middle), well-developed square etch pits are formed; after 35 s (bottom) the as-grown surface features have been erased completely. Vertical lines show the development of some etch pits. The etch pit on the right ends with a flat bottom, which corresponds to the creation of a dislocation during growth. The image area is $1.5 \times 1.5 \mu\text{m}$.

The observation of a high density of growth spirals at the surface of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ films^{3,4} stimulated the idea that dislocations act as strong pinning sites. However, as pointed out by Mannhart *et al.*⁵, only a loose correlation between the number of screw dislocation outcrops and the magnitude of the superconducting current density (j_c) can be established. Moreover, pinning is expected to be about equally strong for screw and edge dislocations⁶. Experiments on miscut⁷ and on bicrystal substrates⁸ indicate the importance of dislocation pinning, without, however, establishing a quantitative relation between the number of defects and the pinning properties. Here we demonstrate that the linear defect structure of a film can be mapped quantitatively, and that an unambiguous relation exists between the as-grown dislocation density n_{disl} and the strong pinning behaviour of j_c in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ thin films.

To determine n_{disl} we etched 140-nm-thick $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ films⁹ in a 1% Br/ethanol solution^{10,11}. The films were pulsed-laser-deposited (PLD) on (100) SrTiO_3 , with a rocking-curve width of the 005-reflection $\Delta\omega < 0.1^\circ$ and a superconducting transition temperature (T_c) of 90 K. Using atomic force microscopy (AFM, Nanoscope III) we find that the typical growth structure consisting of a regularly spaced array of growth islands is erased within ~ 35 s. Square etch pits are formed with a density of 7 to 70 per μm^2 . Etch pits form due to preferential repetitive nucleation of two-dimensional etch islands at those sites where linear defects emerge at the crystal surface. This preferential etching is due to the core energy of the defect and does not depend on the direction of the defect's Burger's vector. Therefore, etch pits are formed both at edge and screw dislocations¹². This point is illustrated by the sputtered film surface shown in Fig. 1a. These films¹³ have the remarkable property that etching does not remove the growth features immediately. It is thereby possible to determine the position of the etch pits with respect to that of the growth spirals typically found on these films. Figure 1a shows a double spiral with a single unit-cell step height which is generated by two screw dislocations with single unit-cell Burger's vectors of opposite sign. The etch pits at the spiral centre prove that such screw dislocations are indeed preferential etch sites. The pits randomly

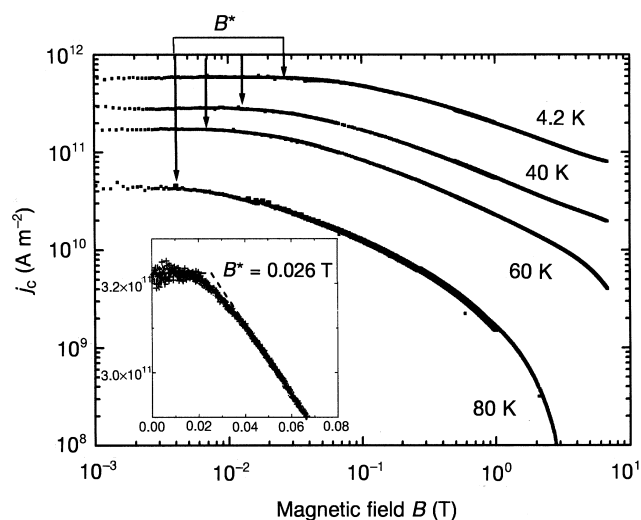


Figure 2 Magnetic field dependence of the superconducting current density j_c of a sample with $n_{\text{disl}} = 6$ per μm^2 at various temperatures. We determine j_c from the magnetic moment of ring patterned films (outer diameter typically 3 mm; width 0.5 mm) by means of a capacitance torque magnetometer¹⁹. At high temperatures, the measured j_c is somewhat reduced by thermal activation²⁰. The inset shows the definition of the characteristic field B^* . In a linear plot of $j_c(B)$, the data are fitted to $j_c(B) = j_c(0)/(1 + k_1 \exp(-1/k_2 x^2))$, with $x = B/k_2$. B^* is determined by the crossing of the tangent at the inflection point and the low-field plateau. (Here α , k_1 and k_2 are fitting constants.)

distributed over the spiral are not related to any growth step, and are therefore due to edge dislocations.

Our PLD films usually grow by two-dimensional nucleation¹⁴. To locate the dislocations in these films, we used a set of engraved markers to reproducibly position the AFM scanner (100- μm scan width) at a well defined area. As the end of a defect is marked by the transition into a flat-bottomed etch pit, we obtain a three-dimensional map of the dislocation structure of a film by following the evolution of an ensemble of etch pits on repeated etching (Fig. 1b). Comparing the as-grown surface with the etched surface, we find that in our PLD films the majority of dislocations (>85%) end at the trenches between the growth islands. The growth process seems to induce a self-organized array of linear defects. On further etching both a high and a low etch pit density film, we find that only ~30% of the etch pits persist from the as-grown film surface down to the substrate interface. This indicates that the majority of linear defects are formed during the growth process. We define the dislocation density n_{disl} as the etch pit density measured after etching for 20 s.

To decide whether linear defects are the main source of strong vortex pinning in thin films, we measured the superconducting current density j_c as a function of temperature ($4.2 < T < 80$ K) and magnetic field ($B \leq 7$ T) for a series of films with various dislocation densities. We found that n_{disl} is proportional to the growth island density (which in turn depends on the substrate temperature and the oxygen pressure during ablation¹⁴), so we can reproducibly 'tune' the defect density by a suitable choice of these growth parameters. In all films, j_c exhibits a plateau for magnetic fields $B < B^*$. Above B^* , a gradual transition to a power-law behaviour—that is, $j_c(B) \propto B^{-\alpha}$ (with $\alpha \approx 0.5$ at low temperatures)—is found (Fig. 2). Plotting the characteristic field B^* at 4.2 K against the measured dislocation density, we find the correlation shown in Fig. 3: B^* is proportional to the dislocation density, and the slope of the fit is $\sim 0.7 \Phi_0$ per dislocation, where Φ_0 is the flux quantum. We conclude that $B^* = 0.7 B_\phi$, where the so-called matching field $B_\phi \equiv n_{\text{disl}} \Phi_0$ corresponds to the density of linear defects naturally present in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ films. We note that no correlation is found between the density of linear defects and the value of the critical current density at $B < B^*$, which shows that the critical current at low fields is solely determined by the pinning of single vortices along extended defects.

This is, as far as we know, the first report of an unambiguous relation between the characteristic field B^* (or the matching field B_ϕ) and the density of natural linear defects in a superconductor. At low magnetic fields, when the number of defects is much larger than the number of vortices, all vortices are pinned and j_c is of the order of the depairing current j_0 (refs 15, 16). For $B < B^*$ we find critical currents up to $7 \times 10^{11} \text{ A m}^{-2}$, close to $j_0 \approx 5 \times 10^{12} \text{ A m}^{-2}$

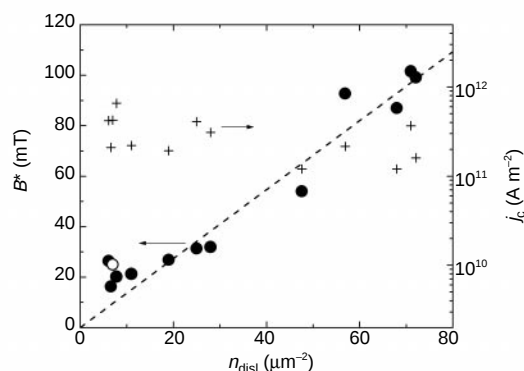


Figure 3 The characteristic field B^* (4.2 K) and the critical current density j_c ($B < B^*$) as a function of the dislocation density n_{disl} . The characteristic field is plotted as filled circles, and the critical current as plus signs; the dislocation density is determined from the measured etch-pit density. The open circle corresponds to the sputtered film shown in Fig. 1a.

(corresponding to a penetration depth $\lambda = 140$ nm and a coherence length $\xi = 1.5$ nm).

At first sight, one might think that the matching behaviour of films is the same as that of single crystals with artificially induced strong pinning defects (for example, columnar defects produced by heavy ion irradiation—see the review by Civile¹⁷). However, there are two essential differences.

First, the nature of the transition in j_c (Fig. 4a). In irradiated crystals, the critical current density above B_ϕ scales as B^{-1} suggesting a collective pinning behaviour; whereas in our films the $B^{-1/2}$ dependence is indicative of a shear of unpinned vortices around those strongly bound to dislocations, albeit with a too-large absolute value for j_c (ref. 5). Below B_ϕ , for films we find a plateau in j_c , whereas in irradiated single crystals j_c decays exponentially up to B_ϕ . According to Krusin-Elbaum *et al.*¹⁸, the exponential decay is due to the inherently random spatial distribution of the artificially produced linear defects in single crystals. In such a case, not all

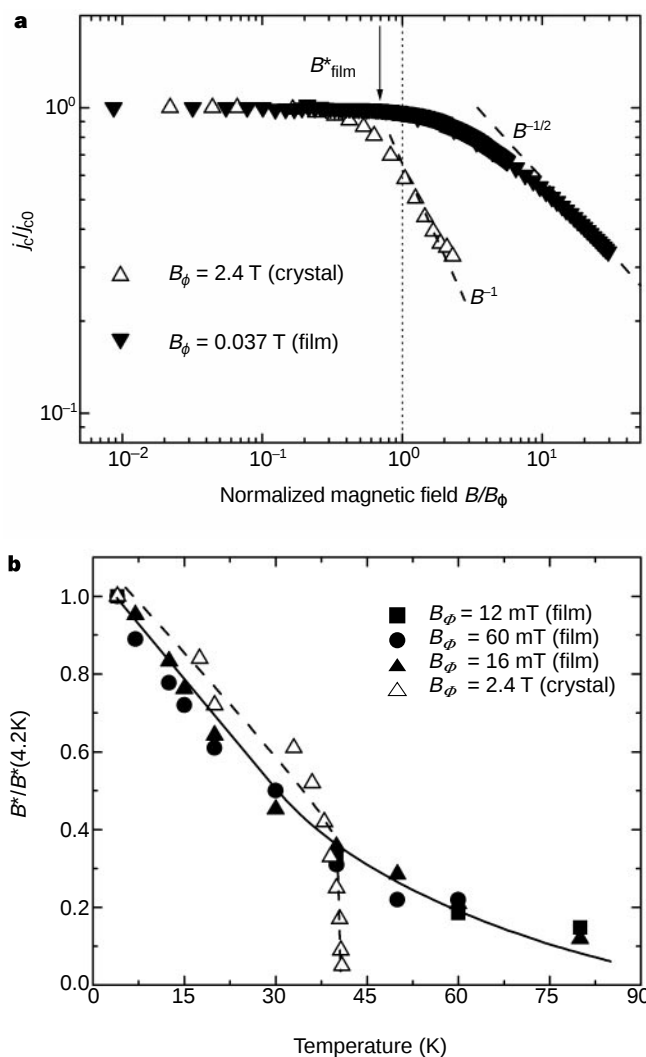


Figure 4 Comparison of a heavy-ion-irradiated $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ single crystal¹⁸ (B_ϕ equals the irradiation dose) with a thin film (B_ϕ equals n_{disl}). **a**, The critical current normalized to j_{c0} , the value of j_c at 4 K. Due to the non-random spatial order of the natural defects, the critical current of the thin films remains approximately constant up to B_ϕ . **b**, The normalized characteristic field B^* . For the thin films, B^* remains finite up to at least 80 K. We note that in the films, the average vortex distance at B_ϕ is about equal to the penetration depth λ . Furthermore, the diameter of dislocations (1–2 nm, ref. 21) is much smaller than that of columnar defects (7–10 nm) and is comparable to the coherence length of the superconductor.

defect sites can accommodate a vortex, as the vortex–vortex interaction dominates over the pinning energy whenever the defects are too close to each other. In contrast, in thin films we find a highly non-random radial distribution function as the growth-induced linear defects are much more evenly distributed. As a result, the vortex lattice easily adapts to the available pinning sites, and the critical current remains constant up to $\sim B_\phi$. The shape of j_c around B_ϕ is determined by the actual radial distribution function, the vortex–vortex interaction and the pinning energy.

Second, the temperature dependence of B^* (Fig. 4b). In thin films we do not observe any deleterious thermal depinning effects, which in irradiated samples reduce B^* to zero at a thermal depinning temperature $T_{dp} = 40$ K. In films, B^* initially also decreases linearly with temperature but, instead of a sudden drop to zero at ~ 40 K, a flattening is observed above 50 K (Fig. 4b). The absence of low-temperature depinning behaviour is in agreement with calculations^{15,16,18} which imply that ideally T_{dp} is above 80 K. The reduced T_{dp} in single crystals is attributed to the fact that the irradiation process not only generates amorphous tracks but also numerous point defects around these tracks¹⁸. Natural linear defects leave the bulk of the superconducting film intact and T_{dp} is not depressed.

To check the robustness of the correlation between B^* and n_{disl} found in PLD films, we also investigated the B^* of sputtered films. In such films, the measured etch pit density (~ 7 per μm^2) is much higher than expected on the basis of the extremely large island size (~ 2.5 - μm diameter). Nevertheless, the value found for B^* follows the trend of PLD films (Fig. 3), which implies that B^* is indeed determined by the density of linear defects and not by some other type of defect related to the island size (for example, the film roughness).

The proportionality between B^* and the dislocation density in both PLD and sputtered films is compelling evidence that the high current densities in $YBa_2Cu_3O_{7-x}$ thin films are due to strong vortex pinning by natural linear defects. This suggests that it may be possible to study vortex matter in films with a tailored non-random distribution of linear strong pinning defects. For technological applications, the challenge is to find ways to increase the density of dislocations while keeping intact their non-random radial distribution. □

Received 1 February; accepted 29 March 1999.

- Hylton, T. L. & Beasley, M. R. Flux-pinning mechanisms in thin films of $YBa_2Cu_3O_{7-x}$. *Phys. Rev. B* **41**, 11669–11672 (1990).
- Douwes, H., Kes, P. H., Gerber, Ch. & Mannhart, J. Critical currents in $YBa_2Cu_3O_{7-x}$ thin films containing screw dislocations. *Cryogenics* **33**, 486–491 (1993).
- Gerber, Ch., Anselmetti, D., Bednorz, J. G., Mannhart, J. & Schlom, D. G. Screw dislocations in high- T_c thin films. *Nature* **350**, 279–280 (1991).
- Hawley, M., Raistrick, I. D., Beery, J. G. & Houlton, R. J. Growth mechanism of sputtered films of $YBa_2Cu_3O_{7-x}$ studied by scanning tunneling microscopy. *Science* **251**, 1587–1589 (1991).
- Mannhart, J. *et al.* Correlation between J_c and screw dislocation density in sputtered $YBa_2Cu_3O_{7-x}$ films. *Z. Phys. B* **86**, 177–181 (1992).
- Pan, V. M., Kasatkin, A. L., Svetchnikov, V. L. & Zandbergen, H. W. Dislocation model of superconducting transport properties of YBCO thin films and single crystals. *Cryogenics* **33**, 21–27 (1993).
- Lowndes, D. H. *et al.* Strong asymmetric flux pinning by miscut-growth-induced columnar defects in epitaxial $YBa_2Cu_3O_{7-x}$ films. *Phys. Rev. Lett.* **74**, 2355–2358 (1995).
- Diaz, A., Mechin, L., Berghuis, P. & Evetts, J. E. Evidence for vortex pinning by dislocations in $YBa_2Cu_3O_{7-x}$ low-angle grain boundaries. *Phys. Rev. Lett.* **80**, 3855–3858 (1998).
- Dam, B. *et al.* The laser ablation threshold of $YBa_2Cu_3O_{7-x}$ as revealed by projection optics. *Appl. Phys. Lett.* **65**, 1581–1583 (1994).
- Vasquez, R. P., Hunt, B. D. & Foote, M. C. Wet chemical techniques for passivation of $YBa_2Cu_3O_{7-x}$. *Appl. Phys. Lett.* **53**, 2692–2694 (1988).
- Roshko, A. *et al.* Effects of etching on the morphology and surface resistance of $YBa_2Cu_3O_{7-x}$ films. *IEEE Trans. Appl. Supercond.* **5**, 1733–1736 (1995).
- Heimann, R. B. *Auslösung von Kristallen* (Springer, Wien, 1975).
- Freisem, S., Brockhoff, A., de Groot, D. G., Dam, B. & Aarts, J. Magnetic & transport properties of sputtered $La_{0.87}Ca_{0.33}MnO_3$ thin films. *J. Magn. Magn. Mater.* **165**, 380–382 (1997).
- Dam, B. *et al.* Growth and etching phenomena on pulsed laser deposited $YBa_2Cu_3O_{7-x}$ films. *Physica C* **261**, 1–11 (1996).
- Blatter, G., Feigel'man, M. V., Gershkenbein, V. B., Larkin, A. I. & Vinokur, V. M. Vortices in high-temperature superconductors. *Rev. Mod. Phys.* **66**, 1125–1388 (1994).
- Nelson, D. R. & Vinokur, V. M. Boson localization and correlated pinning of superconducting vortex arrays. *Phys. Rev. B* **48**, 13060–13097 (1993).
- Civale, L. Vortex pinning and creep in high-temperature superconductors with columnar defects. *Supercond. Sci. Technol.* **10**, A11–A28 (1997).
- Krusin-Elbaum, L., Civale, L., Thomson, J. R. & Field, C. Accommodation of vortices to columnar defects: Evidence for large entropic reduction of vortex localization. *Phys. Rev. B* **53**, 11744–11750 (1996).

- Qvarford, M., Heeck, K., Lensink, J. G., Wijngaarden, R. J. & Griessen, R. Microtorque meter for magnetization measurements on small superconducting samples. *Rev. Sci. Instrum.* **63**, 5726–5732 (1992).
- Schnack, H. G., Griessen, R., Lensink, J. G. & Hai-Hu, W. Generalized inversion scheme for the determination of activation energies from flux-creep experiments on high- T_c superconductors. *Phys. Rev. B* **48**, 13178–13182 (1993).
- Gao, Y., Merkle, K. L., Bai, G., Chang, H. L. M. & Lam, D. J. Structure & composition of grain boundary dislocation cores and stacking faults in MOCVC-grown $YBa_2Cu_3O_{7-x}$ thin films. *Physica C* **174**, 1–10 (1991).

Acknowledgements. This work is part of the research program of FOM (Fundamenteel Onderzoek der Materie) which is financially supported by NWO (Nederlands Wetenschappelijk Onderzoek). A.M.T. thanks the National Research Council of Italy (CNR) for financial support.

Correspondence and requests for materials should be addressed to B.D. (e-mail: dam@nat.vu.nl).

Recovery of surfaces from impurity poisoning during crystal growth

Terry A. Land*, Tracie L. Martin†, Sergey Potapenko*, G. Tayhas Palmore & James J. De Yoreo*

* Department of Chemistry and Materials Science, Lawrence Livermore National Laboratory, Livermore, California 94550, USA

† Department of Chemistry, University of California at Davis, Davis, California 95616, USA

Growth and dissolution of crystal surfaces are central to processes as diverse as pharmaceutical manufacturing^{1,2}, corrosion³, single-crystal production⁴ and mineralization in geochemical and biological environments^{5,6}. Impurities are either unavoidable features of these processes or intentionally introduced to modify the products. Those that act as inhibiting agents induce a so-called 'dead zone', a regime of low supersaturation where growth ceases. Models based on the classic theory of Cabrera and Vermilyea⁷ explain behaviour near the dead zone in terms of the pinning of elementary step motion by impurities^{8,9}. Despite general acceptance of this theory, a number of commonly investigated systems exhibit behaviour not predicted by such models¹⁰. Moreover, no clear microscopic picture of impurity–step interactions currently exists. Here we use atomic force microscopy to investigate the potassium dihydrogen phosphate {100} surface as it emerges from the dead zone. We show that traditional models are not able to account for the behaviour of this system because they consider only elementary steps, whereas it is the propagation of macrosteps (bunches of monolayer steps) that leads to resurrection of growth out of the dead zone. We present a simple physical model of this process that includes macrosteps and relates characteristics of growth near the dead zone to the timescale for impurity adsorption.

Solutions of potassium dihydrogen phosphate (KDP, formula KH_2PO_4) were synthesized from 18-M Ω de-ionized water and high-purity KDP in which the contents of common metal impurities were in the range <50 to 200 parts per billion by weight¹¹. These solutions were stored in Teflon containers to prevent contamination of the solution by aluminium and iron, which are introduced by dissolution of the container wall when stored in typical borosilicate-based glass vessels¹¹. The solute concentration of the solutions was determined by drying to constant weight using standard methods. Supersaturation was defined by $\sigma \equiv \Delta\mu/kT = \ln(C/C_e)$, where $\Delta\mu$ is the change in chemical potential per molecule upon crystallization, k is Boltzmann's constant, T is the absolute temperature, and C and C_e are the actual and equilibrium solute concentrations, respectively. Iron impurities were introduced into these solutions by adding commercial impurity standards consisting of precisely determined levels of Fe^{3+} ions in a dilute nitric acid solution. The supersaturation state was achieved by controlling the temperature of the solution to $\pm 0.1^\circ C$ during the experiments, which were conducted near $33^\circ C$.

Atomic force microscopy was performed on a Digital Nanoscope III. Small (~ 2 -mm) KDP crystals were glued to glass cover slips and placed in the fluid cell of the AFM. *In situ* measurements were obtained in contact mode under conditions of solution flow, using a commercial fluid cell from Digital Instruments. For the solution flow rates used in this study, step speed was independent of flow rate. Thus growth was limited by surface kinetics and not by bulk diffusion. Growth occurred in a step-flow regime on monomolecular steps generated by screw dislocation sources. The monomolecular step height is $3.7 \text{ \AA}$, which corresponds to half the unit cell distance.

The effect of trivalent metal impurities on KDP {100} growth

rates has been investigated in detail using optical techniques such as interferometry^{10,12,13}. Figure 1 shows the dependence of step velocity on supersaturation with the addition of Fe^{3+} impurities. Below a supersaturation given here by σ_d , the crystal exhibits a dead zone where no growth occurs. As the supersaturation is increased beyond σ_d , the surface begins to grow with a step speed that is weakly dependent on σ . Above a sufficiently high supersaturation (denoted here by σ^*), the step velocity rises rapidly, approaching a linear value. Past studies have assumed that, in this linear region, the step velocity gives the speed of the elementary steps, v_e , and is characterized by a kinetic coefficient for elementary steps, β_e , obtained from $v_e = \omega \beta_e \sigma$ where ω is the molecular volume^{10,12}.

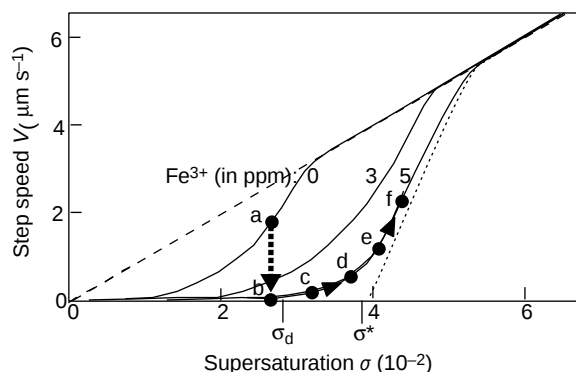


Figure 1 Dependence of step velocity on supersaturation in the presence of Fe^{3+} contamination. Measurements were made interferometrically¹⁰ for the {100} face of KDP. Below a supersaturation given here by σ_d , the crystal exhibits a dead zone where no growth occurs. As the supersaturation increases beyond σ_d , the surface begins to grow. Above a sufficiently high supersaturation (denoted here by σ^*), the step velocity rises rapidly, approaching a linear dependence on supersaturation (dashed line) that is common for all curves and extrapolates to zero. Similar curves are obtained for other impurities, but the detailed shape of the curves

depends on impurity type as well as the purity and pH of the starting solution. The results of our AFM measurements fall very close to these curves (T.A.L., T.L.M., J.J.D. and G.T.P., manuscript in preparation), but for clarity, we have placed points a-f along the curves determined from interferometry, indicating the conditions under which the AFM images of Figs 2 and 3 were collected. The heavy solid lines with arrows show trajectories explored in Figs 2 and 3; the dotted line shows the prediction of standard models for impurity pinning of steps, as discussed in the text.

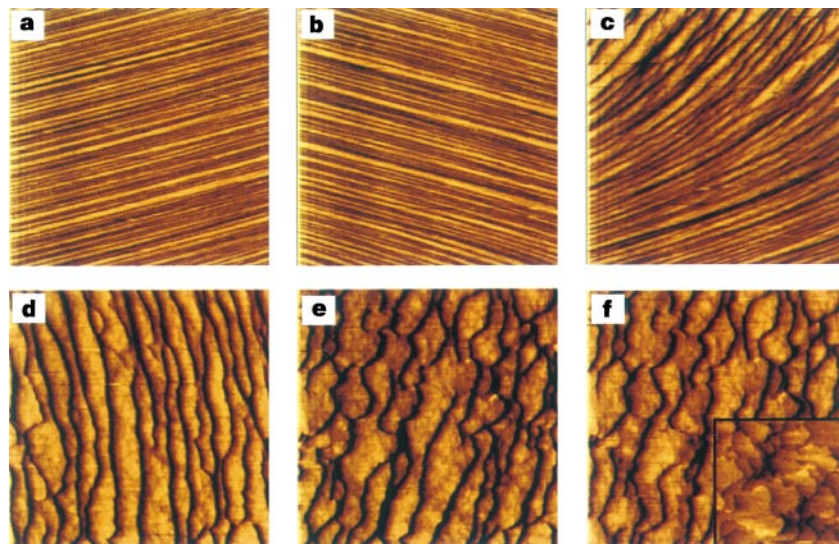


Figure 2 The process of impurity poisoning of the steps at constant supersaturation. These *in situ* AFM images are a sequence of scans (each $15 \times 15 \mu\text{m}$) collected at $\sigma = 0.04$ along the heavy dashed-line trajectory from point a to point b in Fig. 1, showing the poisoning of the surface by introduction of $12 \text{ p.p.m. Fe}^{3+}$ (per mol of KDP). In all images, the fast scan direction is from left to right, and the slow scan direction alternates between images from upwards to downwards. The time at the start of each scan in seconds is: **a**, 0; **b**, 21.3; **c**, 42.7; **d**, 64; **e**, 85.3; and **f**, 106.7. Fe-doped solution was introduced at $t = 42.7 \text{ s}$. Image **c** shows the dramatic slowing of step velocity due to introduction of Fe impurities, and images **d-f** show the subsequent degradation of the surface and near cessation of growth. The

inset to **f** is a $6 \times 6 \mu\text{m}$ image showing the morphology of the elementary steps on the macrostep terraces in **f**. Because steps move during a single scan, the steps have an apparent direction whose angle with the true step direction alternates between negative and positive values for upward and downward scans, respectively. The higher the step speed, the larger the change in angle between up and down scans. The method for obtaining step velocities from these angles is described in ref. 14. Steps that are motionless maintain a constant configuration. In these images and in Fig. 3, the true step direction is nearly vertical with the steps moving from left to right during a scan.

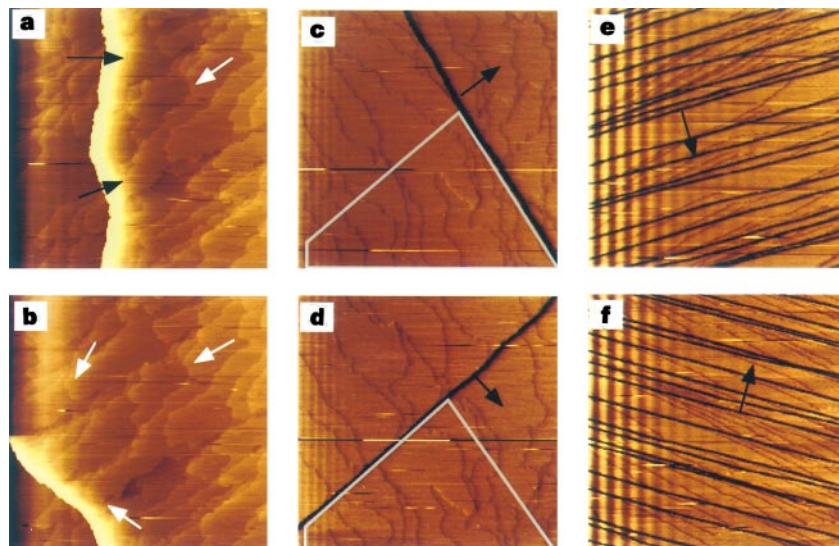


Figure 3 Changes in elementary step and macrostep dynamics with increasing supersaturation. The sequence of images follows the heavy-solid-line trajectory from point b to point f in Fig. 1. The surface morphology at points b and f are illustrated by Fig. 2f and 2a, respectively, while the morphology at points c–e are shown here. The images shown in **a** and **b** (each $9 \times 9 \mu\text{m}$) are downward and upward scans of a macrostep moving from left to right, taken at point c on the curve. Here the elementary steps are practically immobile, and the macrosteps have just begun to move. The images shown in **c** and **d** (each $13 \times 13 \mu\text{m}$) are downward and upward scans collected at point d on the curve. The macrosteps are now straighter and faster but the elementary steps remain nearly immobile.

The macrostep observed in **c** moves out of the image area to the right before the start of **d**. A second macrostep enters the image area from the left shortly after the start of the **d** and advances over part of the surface. During the collection of **c** and **d**, neither macrostep passes through the area outlined by the white polygon. The step configuration within this polygon remains nearly unchanged because the elementary steps are practically immobile. The images shown in **e** and **f** (each $12.5 \times 12.5 \mu\text{m}$) are upward and downward scans taken at point e on the curve. The elementary steps have begun to straighten out and become mobile, but their speed is still considerably less than that of the macrosteps.

In models of the Cabrera–Vermilyea type, impurities adsorbed on the terraces create a field of ‘impurity stoppers’ that act to block the motion of elementary steps. When the average impurity spacing is less than a critical distance, whose magnitude is approximately given by the Gibbs–Thomson critical diameter, the steps can not advance. As the supersaturation is increased and the critical diameter becomes smaller, the steps begin to squeeze through the ‘fence’ of impurities and the step speed rises rapidly. The dotted curve in Fig. 1 illustrates the prediction of such models. No growth occurs below a percolation threshold given by σ^* , while for $\sigma > \sigma^*$, the step speed rises rapidly to its unimpeded value. While the model correctly predicts the rapid rise in step speed for $\sigma > \sigma^*$, no growth is predicted below σ^* , in contrast to what has been observed experimentally.

We have followed the behaviour of steps on the {100} face as a function of supersaturation in the presence of Fe^{3+} impurities using the AFM. Figure 2 illustrates the process of impurity poisoning at constant supersaturation, and follows the trajectory marked by the heavy dashed line with arrows between points a and b in Fig. 1. Figure 2a and b show the initial surface in nominally pure solution. The surface consists of a highly mobile combination of straight elementary steps and step bunches—or macrosteps—moving at constant speed. In this regime, an elementary step has slightly higher speed than a macrostep; this causes step trains, consisting of elementary steps, to gradually decay into a series of macrosteps which consist of about 10 closely spaced elementary steps separated by narrow terraces. (T.A.L., J.J.D., T.L.M. and G.T.P., manuscript in preparation.) The heavy dark lines in this series of images correspond to the macrosteps, while the less common elementary steps are seen as lighter lines. A solution with 12 p.p.m. of Fe^{3+} impurities (mol Fe to mol KDP, equivalent to 5 p.p.m. by weight) was introduced into the flow line. When the Fe^{3+} impurities reached the sample, they had a dramatic effect on step motion and morphology as shown in Fig. 2a and b. As the Fe^{3+} impurities reach the cell, a dramatic slowing of the step velocity occurs. This is

revealed by the bending of the step train seen halfway through this up-scan image. By the end of the first 60 s, both the elementary steps and the macrosteps are distorted and nearly immobile (Fig. 2d). For times greater than 60 s, the surface is in the dead zone. It consists of an intertwined array of highly convoluted macrosteps and elementary steps that are nearly immobile (Fig. 2e and f). The inset to Fig. 2f shows the morphology of the elementary steps in the dead zone. The steps are highly convoluted and exhibit the classic appearance of impurity pinning.

We now follow the changes in step motion and morphology as the supersaturation is increased and the surface is resurrected from the dead zone. This evolution is illustrated in Fig. 3 by three pairs of sequential scans collected at points b–e in Fig. 1. As shown in Fig. 2e and f, when the supersaturation is less than σ_d (point b in Fig. 1), both the single and the macrosteps are distorted and nearly immobile. At supersaturations between σ_d and σ^* (points c and d in Fig. 1), the elementary steps are still nearly immobile and highly convoluted but, as shown in Fig. 3a and b, the macrosteps have just begun to move, the regions immediately behind the macrosteps in each of the two images represent newly formed surfaces, yet they contain many highly ramified elementary steps. But this ramification has not occurred in the commonly assumed fashion by the pinning of advancing elementary steps. Instead, it is a result of impurity-induced separation of elementary steps from the top of the macrostep. The figures show that, as a macrostep advances, it incorporates the immobile elementary steps at its bottom (leading) edge and leaves behind elementary steps that are separated from the top (trailing) edge. It appears that impurities can become attached to the narrow terraces of a macrostep, causing the top step to become pinned in places. While the spacing between impurity sites along some portions of the top step becomes small enough to pin these portions of the step so that they are left behind the advancing macrostep (see white arrows in Fig. 3a and b), the other sections of the step remain part of the macrostep for some period of time (see black arrows). Thus by the time the top step is finally separated from

the macrostep and becomes immobile, it appears ramified and finger-like, as seen in Fig. 2e and f, and Fig. 3a and b. In addition, when a macrostep encounters an immobile elementary step, the latter becomes the bottom step of the macrostep and begins to advance. In this fashion, the height of the macrostep remains roughly constant. By the time point e of Fig. 1 is reached, the macrosteps have straightened out and are moving with increased velocity while the elementary steps still remain nearly immobile (Fig. 3c, d).

These images show that, contrary to the assumptions of the classic impurity pinning models, resurrection out of the dead zone is not a result of elementary steps breaking through the fence of impurities, but rather, it occurs by the motion of macrosteps alone. This observation explains the difference between the experimental curves shown in Fig. 1 and what is expected from the Cabrera–Vermilyea-type models.

At supersaturations near σ^* (point e in Fig. 1), elementary steps begin to straighten out and become mobile, but their speed is still considerably less than that of the macrosteps, as is demonstrated by Fig. 3e, f. In this pair of images, the macrosteps are travelling ~ 1.5 times faster than the elementary steps. Finally, at supersaturations in excess of σ^* (point f in Fig. 1), the surface once again consists of a combination of rapidly moving elementary steps and step bunches that resemble the step trains shown in Fig. 2a and b for the undoped solution. The elementary steps are now moving with a velocity that is close to that of the macrosteps.

The images shown in Fig. 3 demonstrate that the uppermost step in a macrostep is unique, because any event that reduces its speed acts to separate it from the rest of the step bunch. As impurities adsorb onto the terraces, all steps in the macrostep encounter pinning sites. But, as our results show, impurities are only capable of pinning individual elementary steps. Consequently, all impurity pinning sites within the macrostep are buried by the remainder of the macrostep except those along the uppermost step. Because no step follows this one, these pinning sites are permanent.

We have developed a simple model describing the motion of macrosteps, as elementary steps become incorporated and separated from the bunch, that relates σ_d and σ^* to the timescale for impurity adsorption. The model assumes that, following the initial exposure of a terrace, the impurity level on that terrace increases rapidly at first, but then asymptotically approaches its equilibrium value. The characteristic time for approaching this value—that is, the $1/e$ folding time—is given by τ . Because the step speed is a decreasing function of the surface impurity concentration⁷, as impurities adsorb to the terrace in front of this step, there will also be a characteristic time, T , for separation of the uppermost step from the bunch. Analysis (S.Y.P., T.A.L. and J.D.Y., manuscript in preparation) shows that these two times are related by:

$$T = 2\tau\{[n(T)/n_e] + o(v_m/v_e)\} \quad (1)$$

where $n(t)$ is the surface impurity concentration at time t , n_e is its equilibrium value at large t , $o(v_m/v_e)$ is a term of order v_m/v_e which is $\ll 1$, v_m is the speed of the macrostep, and v_e is the speed of an elementary step in the absence of impurities. (The last is given by the dashed line in Fig. 1.)

When the average impurity spacing becomes less than approximately twice the Gibbs–Thomson critical radius, elementary steps become immobile⁷. Because the critical radius is inversely proportional to the supersaturation, the supersaturation at which steps become immobile can be related to surface impurity concentration. For isolated elementary steps, this supersaturation, σ^* , is related to the equilibrium surface impurity concentration, n_e , by $\sigma^* \propto n_e^{0.5}$ (refs 7, 8). In order for elementary steps to be stripped off the macrosteps, the impurity concentration must build up to a value that is sufficient to render the step immobile. When the time, T , for this to occur becomes less than the time for the macrostep to reach the next elementary step, the macrosteps will decay into a network of

intertwined, immobile steps. The surface impurity concentration at time T is then related to the value of σ_d , the supersaturation at which all growth ceases, by $\sigma_d \propto n(T)^{0.5}$. Thus the ratio $n(T)/n_e$ can be replaced by $(\sigma_d/\sigma^*)^2$.

The rate of change in the number of elementary steps, N , in the step bunch is given by:

$$dN/dt = v_m/L - 1/T \quad (2)$$

where L is the average spacing between the elementary steps. The two terms on the right give the rates at which elementary steps are incorporated at the bottom and removed from the top of the macrostep, respectively. Over a small range of supersaturation that lies between σ_d and σ^* , a quasi-steady state is reached where the two terms are balanced and $dN/dt \approx 0$. The images of Fig. 3a–d were obtained in this regime. Combining equations (1) and (2) and ignoring terms of order v_m/v_e gives:

$$\tau = (L/2v_m)(\sigma^*/\sigma_d)^2 \quad (3)$$

Putting the values of L , v_m , σ_d and σ^* into this equation leads to a value for τ , the characteristic time for impurity adsorption, of ~ 10 s. This order of magnitude for τ is consistent with the morphological evolution shown in Figs 2 and 3. If τ were much shorter than this, we would not be able to follow the evolution, given our typical scan time of ~ 20 s. If it were much larger, the observed changes in macrostep structure would only become apparent after many images.

The results presented here explain the anomalous behaviour seen in Fig. 1 where the apparent step velocity rises earlier than is predicted by models of the Cabrera–Vermilyea type. In essence, there are two dead zones: one for macrosteps at $\sigma < \sigma_d$, and another for the elementary steps at $\sigma < \sigma^*$. While the motion of the elementary steps is indeed frozen by $\sigma_d < \sigma < \sigma^*$, macrosteps are still mobile. There is a measurable step velocity, but the apparent kinetic coefficient is that of macrosteps and not elementary steps. As a consequence, the surface morphology, growth rate and degree of impurity incorporation are all determined by the physics of macrosteps. Our results also demonstrate that, in this region, the impurity ions only block the motion of elementary steps. Whereas the inability of individual impurity ions to block the macrosteps is at least conceptually easy to accept, what is less understandable is the ability of the macrosteps to overrun the immobile elementary steps. The best of our knowledge, no model of this process has been proposed. □

Received 6 January; accepted 16 March 1999.

- McPherson, A. *Crystallization of Biological Macromolecules* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999).
- Malkin, A. J. *et al.* Mechanisms of growth for protein and virus crystals. *Nature Struct. Biol.* **2**, 956–959 (1995).
- ASM Handbook Vol. 13, *Corrosion* (eds Davis, J. R. *et al.*) (American Society for Materials, Materials Park, Ohio, 1987).
- Zaitseva, N. P. *et al.* Rapid growth of large-scale (40–55 cm) KH_2PO_4 crystals. *J. Cryst. Growth* **180**, 255–262 (1997).
- Pina, C. M., Becker, U., Risthaus, P., Bosbach, D. & Putnis, A. Molecular-scale mechanisms of crystal growth in barite. *Nature* **395**, 483–486 (1998).
- Teng, H. H., Dove, P. M., Orme, C. A. & De Yoreo, J. J. The thermodynamics of calcite growth: a baseline for biomineralization. *Science* **282**, 724–727 (1998).
- Cabrera, N. & Vermilyea, D. A. *Growth and Perfection of Crystals* 393 (Chapman and Hall, London, 1958).
- Potapenko, S. Yu. Moving of steps through an impurity fence. *J. Cryst. Growth* **133**, 147–154 (1993).
- van Enkevort, W. J. P. & van den Berg, A. C. J. F. Impurity blocking of crystal growth: a Monte Carlo study. *J. Cryst. Growth* **183**, 441–455 (1998).
- Rashkovich, L. N. & Kronsky, N. V. Influence of Fe^{3+} and Al^{3+} ions on the kinetics of steps on the {100} faces of KDP. *J. Cryst. Growth* **182**, 434–441 (1997).
- Zaitseva, N. P., Carman, L., Smolsky, L., Torres, R. & Yan, M. The effect of supersaturation and impurities on the rapid growth of KDP crystals. *J. Cryst. Growth* (in press).
- Rashkovich, L. N. *KDP Family of Crystals* (Adam-Hilger, New York, 1991).
- Dam, B., Bennema, P. & Van Enkevort, W. J. P. In situ observation of surface phenomena on {100} and {101} potassium dihydrogen phosphate crystals. *J. Cryst. Growth* **69**, 306–316 (1986).
- Land, T. A., De Yoreo, J. J. & Lee, J. D. An *in situ* AFM investigation of canavalin crystallization kinetics. *Surf. Sci.* **384**, 136–155 (1997).

Acknowledgements. We thank L. Carman for providing the starting solutions, R. Torres for chemical analyses, P. Rock and W. Casey for use of equipment, and the NSF for support. This work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory.

Correspondence and requests for materials should be addressed to T.A.L. (e-mail: land1@llnl.gov).

Tying a molecular knot with optical tweezers

Yasuharu Arai^{*†}, Ryohei Yasuda^{*†}, Ken-ichirou Akashi^{*}, Yoshie Harada^{*†}, Hidetake Miyata^{*‡}, Kazuhiko Kinoshita Jr^{*†} & Hiroyasu Itoh^{§†}

^{*} Department of Physics, Faculty of Science and Technology, Keio University, Hirosaki 3-14-1, Kohoku-ku, Yokohama 223-8522, Japan

[†] CREST (Core Research for Evolutional Science and Technology) "Genetic Programming" Team 13, Nogawa 907, Miyamae-ku, Kawasaki 216-0001, Japan

[§] Tsukuba Research Laboratory, Hamamatsu Photonics KK, Tokodai, Tsukuba 300-2635, Japan

Filamentous structures are abundant in cells. Relatively rigid filaments, such as microtubules and actin, serve as intracellular scaffolds that support movement and force, and their mechanical properties are crucial to their function in the cell. Some aspects of the behaviour of DNA, meanwhile, depend critically on its flexibility—for example, DNA-binding proteins can induce sharp bends in the helix¹. The mechanical characterization of such filaments has generally been conducted without controlling the filament shape, by the observation of thermal motions^{2–5} or of the response to external forces^{6–9} or flows^{10–12}. Controlled buckling of a microtubule has been reported¹³, but the analysis of the buckled shape was complicated. Here we report the continuous control of the radius of curvature of a molecular strand by tying a knot in it, using optical tweezers to manipulate the strand's ends. We find that actin filaments break at the knot when the knot diameter falls below 0.4 μm . The pulling force at breakage is around 1 pN, two orders of magnitude smaller than the tensile stress of a straight filament. The flexural rigidity of the filament remained unchanged down to this diameter. We have also knotted a single DNA molecule, opening up the possibility of studying curvature-dependent interactions with associated proteins. We find that the knotted DNA is stronger than actin.

To manipulate an actin filament, two polystyrene beads coated with myosin were held with dual-beam optical tweezers¹⁴. A fluorescent actin filament was moved toward the beads, by moving the sample stage; one end of the filament was attached to a bead, and then the other end to the other bead. Subsequently, one end was manipulated with the optical tweezers, as shown in Fig. 1, to form a knot. To facilitate the manipulation, the medium contained sucrose (50% by weight; 50 wt%) which suppressed the brownian motion. When the end was to cross the filament (Fig. 1, images 3, 5 and 7), the end beads were brought either up or down by changing the microscope focus while the remainder of the filament stayed at the same height in the medium. The process of knotting took <1 min. Note that the transparent grip by the optical tweezers allowed continuous manipulation during knotting, without a shift of the grip as would be required in knotting with fingers or mechanical tweezers.

Once knotted, the filament was pulled stepwise to reduce the diameter of the knot and, thus, to establish the relation between the tension and knot diameter. In Fig. 2a, the upper bead was moved upwards (Fig. 2b), while the lower bead was held weakly in a stationary trap. The tension on the filament was estimated from the displacement of the lower bead from the centre of the optical trap (Fig. 2c). Under a tension appreciably above the noise level, the knot diameter was too small to be resolved in the microscope image, so we estimated it in two ways. (1) We compared the intensity of the knotted region with that of a straight region (Fig. 2a). (2) We

measured the knot's diameter in the image when the diameter was 0.6–0.8 μm . Starting from this reference point, further increase in the distance between the end beads was equated to the decrease in the circumference of the knot (after correction for the apparent increase in the filament length at high tensions due to suppression of the brownian fluctuation, the amount of which was estimated in unknotted filaments). The two methods gave consistent results, although the latter method was subject to a larger error at the reference point.

In Fig. 2d, the tension (T) is plotted against the knot diameter (D) estimated from the fluorescence intensity. Data from 26 knotted filaments are combined. Although the scatter is large, the data are consistent with the smooth solid line representing the relation $T = 2\kappa/D^2$ expected for a homogeneous rod with a flexural rigidity κ . (The elastic energy U of a ring of diameter D formed from a straight cylinder is given¹⁵ by $U = 2\pi EI/D$, where E is the Young's modulus and I the moment of inertia of the rod ($EI = \kappa$). Pulling a knotted cylinder by dL reduces the knot diameter by $dD = dL/\pi$, and thus the tension required is $T = -dU/dL = 2\kappa/D^2$.) The flexural rigidity calculated from this equation was $(5.5 \pm 0.2) \times 10^{-26} \text{ N m}^2$ (mean $\pm$ standard error) over the plotted diameter range. This value is close to the rigidity reported for free actin filaments undergoing brownian motion^{3–5} of $(5.8\text{--}7.4) \times 10^{-26} \text{ N m}^2$, which corresponds to a persistence length ($= \kappa/k_B T_A$, where k_B is the Boltzmann constant and T_A the absolute temperature) of 14–18 μm (an early estimate¹⁶ was $\sim 6 \mu\text{m}$). Thus, an actin filament bent to a radius of only 0.2 μm is as flexible as a free filament for which the radius of curvature induced by brownian motion is, on the average, tens of micrometres. An indirect estimate of κ for bent actin also supports this notion⁸.

When the knot diameter became $\sim 0.36 \mu\text{m}$, or when the tension in the filament became $\sim 0.9 \text{ pN}$, the filament broke at the knot (Figs 2a, 3a, 3b). This point of breakage for a knotted polymer is predicted theoretically¹⁷. The fracture occurred, in most filaments, within 10 s after the tension was raised to $\sim 1 \text{ pN}$. The time elapsed before the fracture tended to be shorter for a higher tension, as in the case of unbinding of actin from myosin¹⁸ or α -actinin¹⁹, but the scatter in the present data precluded precise analysis of the time dependence. We were unable to break an unknotted actin filament: the filament was detached from a myosin-coated bead at a tension between 3–14 pN, consistent with the unbinding force between myosin and actin¹⁸ of $\sim 10 \text{ pN}$. The tensile strength of straight actin filaments is⁹ $\sim 600 \text{ pN}$. An actin filament thus resists straight pull

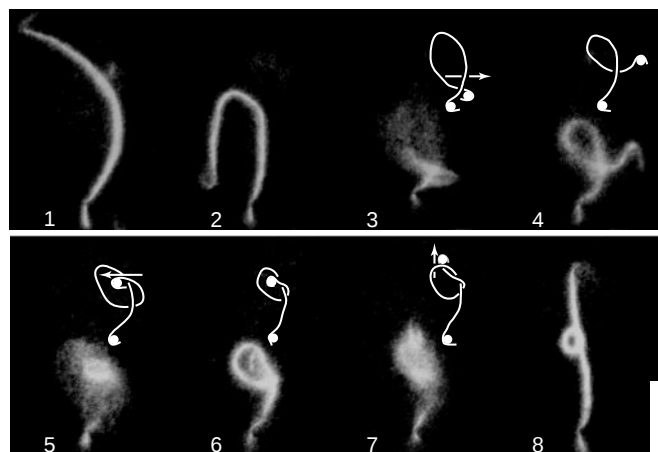


Figure 1 Tying a knot in an actin filament. Explanatory drawings are added in images 3–7. In images 3 and 7, the microscope focus was moved to below the filament to bring down the end beads which were trapped at the focus level; the focus was above the filament in image 5. Scale bar, 10 μm .

[‡] Present address: Physics Department, Graduate School, Tohoku University, Aoba-ku, Sendai 980-8578, Japan.

but easily yields to bending. The critical tension of ~ 1 pN for knotted actin did not change appreciably in three sets of conditions that potentially increase the stability of an actin filament: an increase in the KCl concentration to 100 mmol per litre of the solution containing 50% sucrose, the addition of MgCl_2 at 5 mmol per litre, or the incubation with ~ 60 -fold molar excess of phalloidin for 8 h. Also, changing the duration of fluorescence excitation, which might cause photo-damage, between 2 and 10 min did not affect the results. Thus, the small fracture strength of knotted and bent filaments is unlikely to be an experimental artefact. Twisting, too, greatly reduces the fracture strength of actin filaments⁹: 10- μm

filaments twisted by only one turn in either direction gave a tensile strength of ~ 200 pN, a threefold reduction from the strength of straight actin.

At the knot diameter of $0.36 \mu\text{m}$ (Fig. 3c), the amount of shear per actin monomer is ~ 0.15 nm at the filament radius of 2.5 nm, half the maximal radius of an actin filament²⁰. This relatively small shear, about half the interatomic distances, appears to be sufficient to break the bonds between actin monomers on the timescale of seconds. If an actin filament is modelled as a homogeneous rod with a flexural rigidity of $5.5 \times 10^{-26} \text{ N m}^2$, a lateral pull of only 2.8 pN at the centre between two supports separated by $0.35 \mu\text{m}$ (Fig. 3d) would produce a curvature with a radius of $0.18 \mu\text{m}$. Even a single myosin molecule can produce this much force²¹. Actin filaments in a cell may be easily broken if forces act perpendicularly on them; for example, through α -actinin¹⁹. (Note that the rigid supports in Fig. 3d were introduced for the purpose of calculation. The supports exert a force of 1.6 pN in the direction perpendicular to the filament. The combination of these two forces with the pull of 2.8 pN at the top is sufficient to produce the radius of curvature of $0.18 \mu\text{m}$.)

A single molecule of DNA has been knotted similarly (Fig. 4); even multiple knots could be introduced in a single DNA molecule (not shown). To facilitate knotting, unstrained actin filaments were added to the medium at $10 \mu\text{mol kg}^{-1}$ in addition to 46 wt% sucrose (unstrained DNA may also serve this purpose, because inadvertent knotting was observed in a thick solution of DNA¹⁰). Thus, the knot encircled many actin filaments, and the filaments were bundled when the DNA was pulled taut. This is evident in Fig. 4, where the knotted DNA was first pulled upwards (image 15) and released (image 16), then pulled downwards (image 17) and released (image 18). The knot position remained unchanged during the total of three such cycles (the additional cycles are not shown), indicating that bundled filaments in the knot prevented the knot from moving. We have been unable to break knotted or unknotted DNA with the optical tweezers. One of the beads escaped from the optical trap, as shown in Fig. 4 (images 4–5 and 21–24), indicating that the fracture strength was greater than 15 pN. Unknotted DNA has been reported to transform into an overstretched state^{6,7} at 65–70 pN and to break¹¹ at 476 pN.

At present, knotting takes tens of seconds and hence requires a

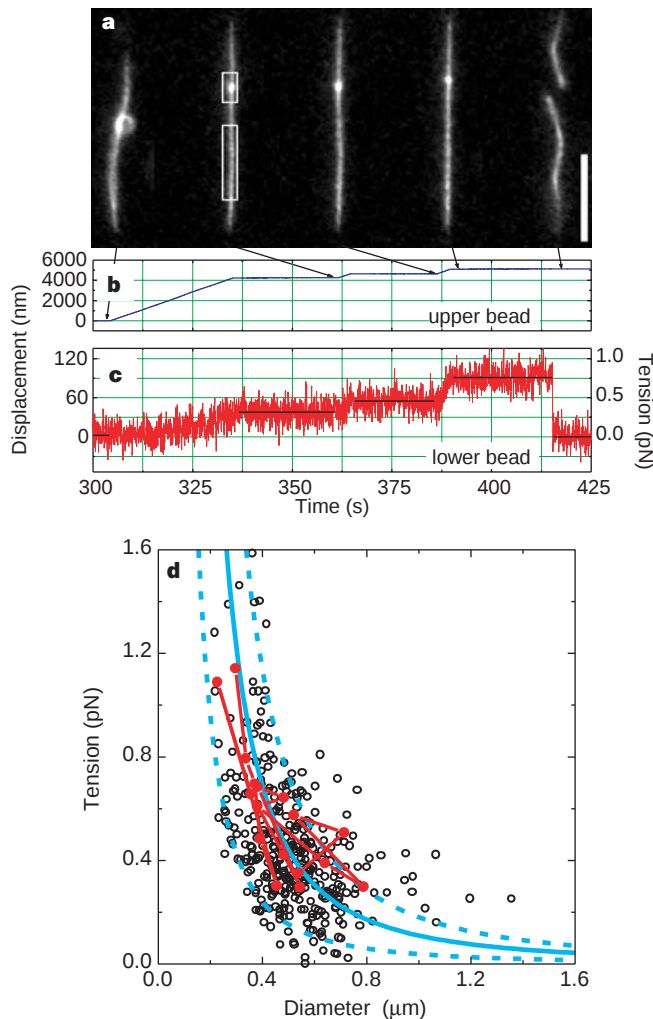


Figure 2 Relation between the tension (T) and knot diameter (D). **a**, Fluorescence images of the filament. D was estimated from the background corrected intensities per pixel, I_{knot} and I_{straight} , averaged over respective rectangles: $D = (I_{\text{knot}} - I_{\text{straight}})L_{\text{knot}}/\pi I_{\text{straight}}$ where L_{knot} is the length of the rectangle enclosing the knot portion. Scale bar, $10 \mu\text{m}$. **b**, Movement of the upper bead trapped at a stiffness of 0.086 pN nm^{-1} . **c**, Displacement of the lower bead in a trap at $0.0086 \text{ pN nm}^{-1}$. T equals the displacement times the stiffness. The filament broke at the knot at 415 s. The abscissa is the time elapsed after knotting, during which several pull-release cycles were made. **d**, Tension-diameter relationship for 26 actin filaments. Data for one filament are connected with red lines, in the order of observation, to demonstrate that the knot diameter changed reversibly in response to the tension change. Blue lines represent $T = 2\kappa/D^2$ for $\kappa = (5.5 \pm 3.6) \times 10^{-26} \text{ N m}^2$; solid line, the average; dashed lines, standard deviations for the 356 data points. Estimated uncertainty for each point is ± 0.2 pN in T (mainly due to drift in the trap position) and $\pm 0.2 \mu\text{m}$ in D (from the differences in D values obtained by the two methods explained in text).

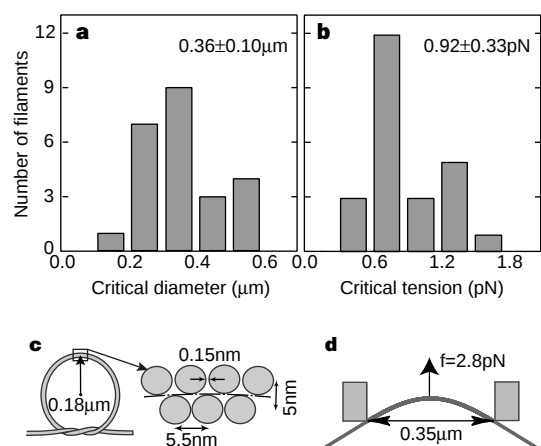


Figure 3 Breakage parameters of a knotted actin filament. **a**, Critical knot diameter; **b**, tension at breakage. The average $\pm$ standard deviation are shown ($n = 24$). **c**, A highly schematic diagram of an actin filament bent to a radius of curvature, r , of $0.18 \mu\text{m}$. Ellipsoids represent actin monomers. **d**, Force f required to bend a rod supported at two points¹⁵. For the support separation L_0 of $0.35 \mu\text{m}$, $f = 2.8$ pN produces $r = 0.18 \mu\text{m}$ at the pulled point. (After correcting typographic errors in ref. 15, we obtained a relation $f = (EI/r^2)\tan\theta_0$, where θ_0 is a parameter (the angle between f and the filament at the support edge) determined by $L_0/r = \int_0^{\pi/2} d\theta [2(\cos\theta_0/\cos\theta)^{1/2} \cos(\theta - \theta_0)]$.)

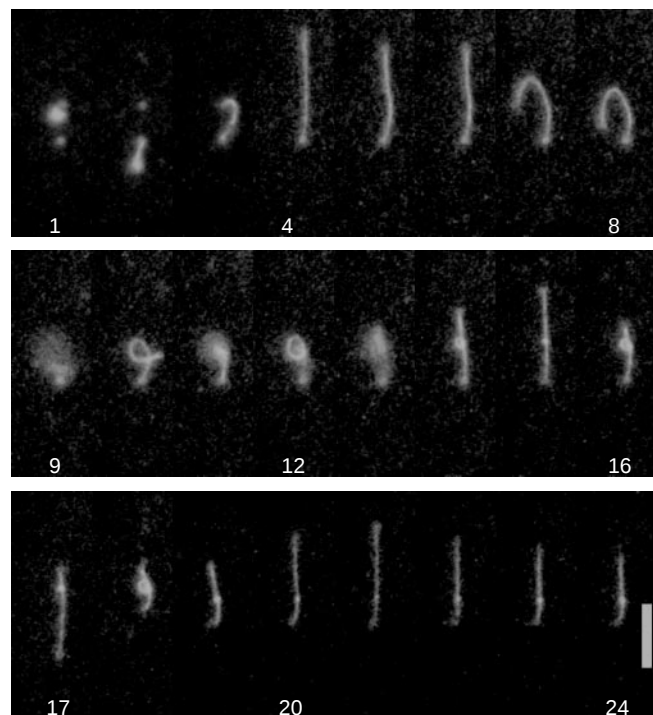


Figure 4 Knotting a single DNA molecule. Two beads were held with optical tweezers, and a DNA molecule stained with POPO-3 was attached to the lower bead by moving the microscope stage (images 1 and 2). The stage was moved upwards to pull the DNA, and the other end was attached to the upper bead (image 3). The upper bead was then moved with the optical tweezers to straighten the DNA (4) and to confirm firm attachment; the upper bead escaped from the trap when the upper trap went too far (4 and 5). After re-trapping, a knot was made as in Fig. 1a (6–14). The knot diameter was made smaller and larger by moving the upper bead up (15) and down (16), or moving the lower bead down (17) and up (18). After two more cycles (not shown), the upper bead was brought up until it escaped from the trap without rupture in the DNA (21–24). The knot diameter in image 21 is estimated as $<0.2 \mu\text{m}$ from its intensity. Scale bar, $10 \mu\text{m}$.

medium with a high effective viscosity which suppresses the brownian motion. However, rapid knotting should be possible by electronically programming the movement of the optical tweezers (and the microscope stage). If the viscosity was made low enough, it should be possible to introduce, after knotting, a different solution containing, for example, a protein that would interact with the filament. At least two applications of our technique are conceivable, other than its use for the mechanical measurements described here. One is to study the curvature dependence of the interaction of intracellular filaments with associated proteins or ligands. For example, RNA polymerase has been shown to be released from DNA when the DNA is pulled straight²². Measurements of association constants at precisely controlled curvatures are desirable; such measurements are applicable to the study of the dynamics of cytoskeletal networks, where many protein molecules interact with cytoskeletal filaments. The other possibility is to use a knot as a manipulation tool, as in the bundling of actin filaments by DNA. It might be possible to constrict a cell into halves by tying a knot round it, for mimicking cytokinesis or for artificial compartmentalization; or one could constrict a cell protrusion such as a dendrite. The force from optical tweezers may not be strong enough for such operations, but it may be possible to make three knots in a single DNA molecule and, using the two knots at the ends, fasten the DNA string to two microneedles (coated with suitable glue) attached to manipulators. □

Methods

Materials. Rabbit skeletal actin and myosin were prepared²³ and stained²⁴ with phalloidin-tetramethylrhodamine B isothiocyanate conjugate (Fluka). Amino-polystyrene beads ($0.9 \mu\text{m}$ diameter, Polysciences) were covalently bound with myosin⁸. To tie an actin knot, a solution containing, per kg, 500 g of sucrose, 7 mmol imidazole (pH 7.4), 7 mmol KCl, 3 mmol 2-mercaptoethanol, 5 g glucose, 0.3 g glucose oxidase, 0.07 g catalase, 7 mg myosin-coated beads, and 0.15 nmol labelled F-actin was put into an observation chamber²⁴ pretreated with a solution of BSA (10 mg ml^{-1}). λ -phage DNA was biotinylated at both ends²² and attached to streptavidin-coated beads²² ($0.95 \mu\text{m}$) in a solution containing, per kg, 460 g sucrose, 6 mmol HEPES (pH 7.4), 17 mmol KCl, 0.6 mmol EDTA, 7 mmol 2-mercaptoethanol, 2 g glucose, 0.1 g glucose oxidase, 0.02 g catalase, 7 mg streptavidin-coated beads, 0.1 g α -casein, 13 μmol F-actin, 7 pmol DNA, and 70 nmol POPO-3 (Molecular Probes).

Microscopy. Actin filaments and DNA were knotted using an inverted fluorescence microscope (Diaphot TMD, Nikon) equipped with dual optical tweezers and two video cameras for simultaneous observation of phase-contrast images of beads and fluorescence images of actin filaments or DNA¹⁴. The trap stiffness of the tweezers was measured after each experiment by trapping a free bead at a reduced laser intensity and observing the width of the brownian motion within the trap¹⁴; correction for the finite temporal resolution of the video camera was not necessary because the bead motion was slow in the viscous media. The position of a bead was measured as the centroid of the phase-contrast image of the bead. Temperature, $23 \pm 3^\circ\text{C}$.

Received 8 February; accepted 13 April 1999.

- Schultz, S. C., Shields, G. C. & Steitz, T. A. Crystal structure of a CAP-DNA complex: the DNA is bent by 90° . *Science* **253**, 1001–1007 (1991).
- Yanagida, M., Hiraoka, Y. & Katsura, I. Dynamic behaviors of DNA molecules in solution studied by fluorescence microscopy. *Cold Spring Harbor Symp. Quant. Biol.* **47**, 177–187 (1983).
- Yanagida, T., Nakase, M., Nishiyama, K. & Oosawa, F. Direct observation of motion of single F-actin filaments in the presence of myosin. *Nature* **307**, 58–60 (1984).
- Gittes, F., Mickey, B., Nettleton, J. & Howard, J. Flexural rigidity of microtubules and actin filaments measured from thermal fluctuations in shape. *J. Cell Biol.* **120**, 923–934 (1993).
- Yasuda, R., Miyata, H. & Kinosita, K. Jr Direct measurement of the torsional rigidity of single actin filaments. *J. Mol. Biol.* **263**, 227–236 (1996).
- Cluzel, P. et al. DNA: an extensible molecule. *Science* **271**, 792–794 (1996).
- Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271**, 795–799 (1996).
- Dupuis, D. E., Guilford, W. H., Wu, J. & Warshaw, D. M. Actin filament mechanics in the laser trap. *J. Muscle Res. Cell Motil.* **18**, 17–30 (1997).
- Tsuda, Y., Yasutake, H., Ishijima, A. & Yanagida, T. Torsional rigidity of single actin filaments and actin-actin bond breaking force under torsion measured directly by *in vitro* micromanipulation. *Proc. Natl Acad. Sci. USA* **93**, 12937–12942 (1996).
- Perkins, T. T., Quake, S. R., Smith, D. E. & Chu, S. Relaxation of a single DNA molecule observed by optical microscopy. *Science* **264**, 822–826 (1994).
- Bensimon, D., Simon, A. J., Croquette, V. & Bensimon, A. Stretching DNA with a receding meniscus: experiments and models. *Phys. Rev. Lett.* **74**, 4754–4757 (1995).
- Smith, D. E. & Chu, S. Response of flexible polymers to a sudden elongational flow. *Science* **281**, 1335–1340 (1998).
- Kurachi, M., Hoshi, M. & Tashiro, H. Buckling of a single microtubule by optical trapping forces: direct measurement of microtubule rigidity. *Cell Motil. Cytoskel.* **30**, 221–228 (1995).
- Miyata, H. et al. Mechanical measurements of single actomyosin motor force. *Biophys. J.* **68**, 286S–290S (1995).
- Landau, L. D. & Lifshitz, E. M. *Theory of Elasticity* 2nd edn (Pergamon, Oxford, 1970).
- Oosawa, F., Fujime, S., Ishiwata, S. & Mihashi, K. Dynamic property of F-actin and thin filament. *Cold Spring Harbor Symp. Quant. Biol.* **37**, 277–285 (1973).
- Saitta, A. M., Soper, P. D., Wasserman, E. & Klein, M. L. Influence of a knot on the strength of a polymer strand. *Nature* **399**, 46–48 (1999).
- Nishizaka, T., Miyata, H., Yoshikawa, H., Ishiwata, S. & Kinosita, K. Jr Unbinding force of a single motor molecule of muscle measured using optical tweezers. *Nature* **377**, 251–254 (1995).
- Miyata, H., Yasuda, R. & Kinosita, K. Jr Strength and lifetime of the bond between actin and skeletal muscle α -actinin studied with an optical trapping technique. *Biochim. Biophys. Acta* **1290**, 83–88 (1996).
- Holmes, K. C., Popp, D., Gebhard, W. & Kabsch, W. Atomic model of the actin filament. *Nature* **347**, 44–49 (1990).
- Finer, J. T., Simmons, R. M. & Spudis, J. A. Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature* **368**, 113–119 (1994).
- Harada, Y. et al. Single-molecule imaging of RNA polymerase-DNA interactions in real time. *Biophys. J.* **76**, 709–715 (1999).
- Suzuki, N., Miyata, H., Ishiwata, S. & Kinosita, K. Jr Preparation of bead-tailed actin filaments: estimation of the torque produced by the sliding force in an *in vitro* motility assay. *Biophys. J.* **70**, 401–408 (1996).
- Yasuda, R., Noji, H., Kinosita, K. Jr & Yoshida, M. F₁-ATPase is a highly efficient molecular motor that rotates with discrete 120° steps. *Cell* **93**, 1117–1124 (1998).

Acknowledgements. We thank K. Adachi and Y. Hada for technical support and advice, S. Ishiwata for critically reading the manuscript, M. Hosoda for image processing, and H. Umezawa for laboratory management. This work was supported in part by Grants-in-Aid from Ministry of Education, Science, Sports and Culture of Japan, an Academic Frontier Promotional Project, and a Keio University Special Grant-in-Aid. R.Y. was a Research Fellow of the JSPS.

Correspondence and requests for materials should be addressed to Y.H. at Keio University.

Memory of macromolecular helicity assisted by interaction with achiral small molecules

Eiji Yashima^{*†}, Katsuhiko Maeda[†] & Yoshio Okamoto[‡]

^{*} Form and Function, Precursory Research for Embryonic Science and Technology, Japan Science and Technology Corporation, [†] Department of Molecular Design and Engineering, [‡] Department of Applied Chemistry, Graduate School of Engineering, Nagoya University, Chikusa-ku, Nagoya 464-8603, Japan

The helicity of biological macromolecules such as DNA and proteins is largely governed by the homochirality of their components (D-sugars and L-amino acids). In polymer and supramolecular chemistry, control of helicity is an attractive goal because of possible applications in materials science, chemical sensing and enantioselective catalysis^{1–13}. We reported recently that macromolecular helicity can be induced in a polymer by an optically active amine¹⁴. Here we show that this helicity can be ‘memorized’ when the amine is replaced by various achiral amines. Although the maintenance of helicity in the polymer is not perfect, it can ‘repair’ itself over time. Small structural changes in the achiral amines influence the efficiency of helicity retention markedly.

We recently reported the induction of helicity in a stereoregular, *cis-transoidal* poly((4-carboxyphenyl)acetylene) (poly-1; Fig. 1)¹⁴. The polymer possesses a large number of short helical units with many helix-reversal points, and is therefore achiral. However, in the

presence of optically active amines such as (R)-2 and (S)-3 capable of interacting with the polymer's carboxy groups, a dynamic, one-handed macromolecular helicity is induced in the polymer, resulting in optical activity (Fig. 1). Analogous induction of macromolecular helicity occurs in the tobacco mosaic virus, which induces a right-handed helix in RNA on the inside of the protein shell¹⁵. The complexes of poly-1 with optically active amines show a characteristic, induced circular dichroism (ICD) in the ultraviolet–visible region which depends on the absolute configuration of the chiral amines, and the signs corresponding to the helix-sense can be used as a probe for the chirality assignment of the amines. We now find that the macromolecular helicity of poly-1 induced by optically active amines remains even after removal and replacement of the optically active amines by achiral ones (Fig. 1).

Figure 2, trace a shows the circular dichroism (CD) spectrum of poly-1 in the presence of (R)-2 (10 equiv. to monomer units of poly-1) in dimethyl sulphoxide (DMSO). The poly-1–(R)-2 complex shows a characteristic, exciton-type coupled ICD¹⁶. As expected, the optical activity of the poly-1–(R)-2 complex instantly disappeared when the complex was exposed to a stronger acid (such as trifluoroacetic acid), which frees the poly-1 so it reverts to the original, optically inactive (non-helical) polymer. On addition of 5 equiv. (S)-3 to the solution of poly-1–(R)-2, the ICD spontaneously changed and the signs inverted to give mirror images (Fig. 2, trace b). This clearly indicates that the induced helix of the poly-1 is dynamic in nature, and the helix-sense can be controlled with a rather small amount of the chiral amino alcohol (S)-3 (Fig. 1) with the opposite configuration; we know this because an acid–base binding constant (*K*) of (S)-3 with 4-vinylbenzoic acid (a model compound of poly-1), obtained by standard ¹H NMR titration experiments¹⁴ in dry DMSO-*d*₆, is larger (509 ± 30 M^{–1}) than that

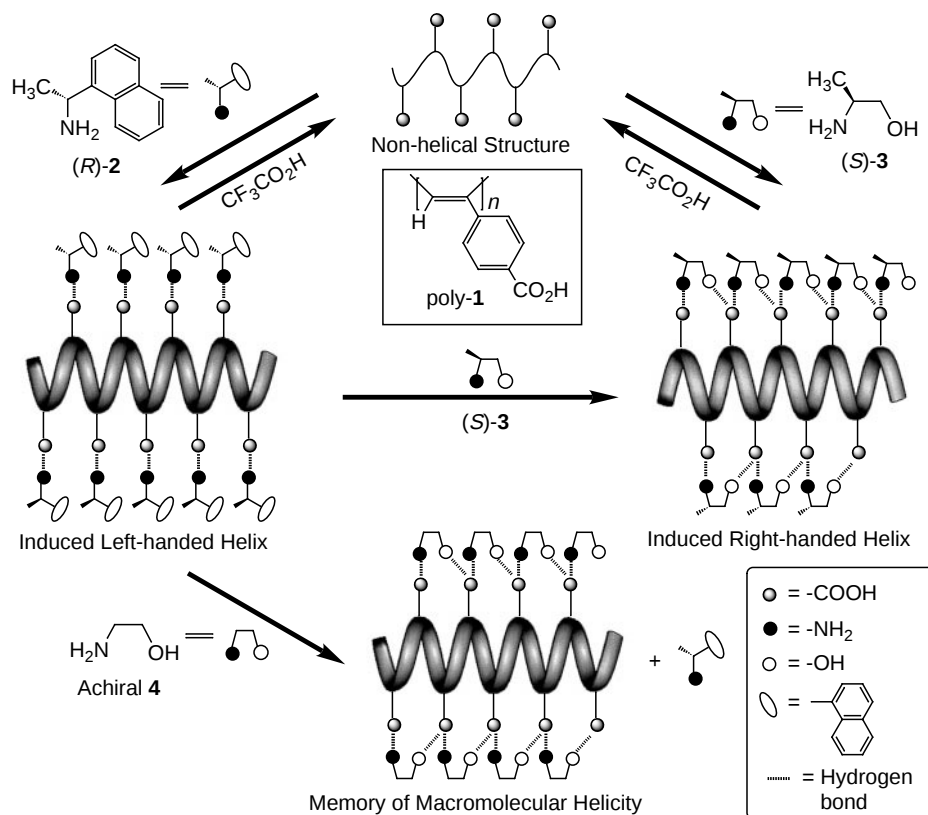


Figure 1 Schematic illustration of induced one-handed helicity in poly-1, and the memory of macromolecular helicity. A one-handed helical conformation of poly-1 is induced with (R)-2 (middle left) or (S)-3 (middle right) and the induced macromolecular helicity is memorized by achiral 4 (bottom). The helix-senses of poly-1 induced by (R)-2 and (S)-3 are tentative, but they should be opposite, as the

induced CD signs corresponding to the helix-sense of poly-1 depend on the absolute configuration of the chiral amines¹⁴. The systematic names of 2, 3 and 4 are (R)-1-(1-naphthyl)ethylamine, (S)-2-amino-1-propanol and 2-aminoethanol, respectively.

of (*R*)-**2** ($57 \pm 2 \text{ M}^{-1}$) and, therefore, the (*R*)-**2** is completely replaced by the (*S*)-**3** during the titration.

We then decided to use an achiral amino alcohol, 2-aminoethanol (**4**, 50 equiv.; see Fig. 1), instead of (*S*)-**3** and added it to the poly-**1**–(*R*)-**2** complex. Compound **4** is a strong base similar to (*S*)-**3** ($K = 1,050 \pm 160 \text{ M}^{-1}$), so the (*R*)-**2** complexed with poly-**1** will be completely replaced by an excess of **4**. However, rather surprisingly, the ICD signal was still observed, with only a slight decrease in intensity (Fig. 2, trace c). This indicates that the one-handed helical conformation of the poly-**1** induced by (*R*)-**2** may be memorized even after the (*R*)-**2** is replaced by achiral **4**. To investigate this process further, we isolated poly-**1** from the poly-**1**–(*R*)-**2** complex by gel permeation chromatography (GPC)

using a DMSO solution of **4** (0.8 M) as the mobile phase (see Methods). The poly-**1** and (*R*)-**2** were completely separated and fractionated by GPC. Based on the ultraviolet (UV) spectrum of the (*R*)-**2** fraction, more than 99% of the (*R*)-**2** was recovered. Nevertheless, the poly-**1** fraction (0.13 mg ml^{-1}) containing a large excess of achiral **4** (~ 900 equiv.) showed an intense ICD, comparable to that measured before the GPC fractionation (Fig. 2, trace d). The fractionated, optically active poly-**1** was injected again into the GPC system using the same mobile phase. However, we could not detect any trace amounts of (*R*)-**2**, and the re-fractionated poly-**1** exhibited ICD without any loss of the macromolecular helicity. We further repeated this procedure, but no change in the CD spectrum of the poly-**1** was observed. Moreover, we found that the addition of

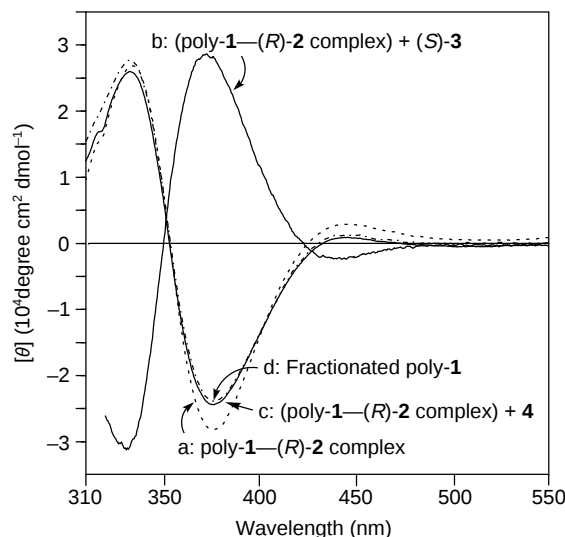


Figure 2 Changes of the CD spectra of poly-**1** with amines. Shown are the CD spectra of the poly-**1**–(*R*)-**2** complex (trace a), the mixture of the poly-**1**–(*R*)-**2** complex with (*S*)-**3** (trace b) and **4** (trace c), and the isolated poly-**1** (trace d)

obtained by GPC using a DMSO solution of **4** (0.8 M) as the mobile phase, in DMSO at ambient temperature (20–25 °C). The concentration of poly-**1** is $3.0 \text{ mg (20.4 } \mu\text{mol monomer units) ml}^{-1}$ for traces a–c and 0.13 mg ml^{-1} for trace d.

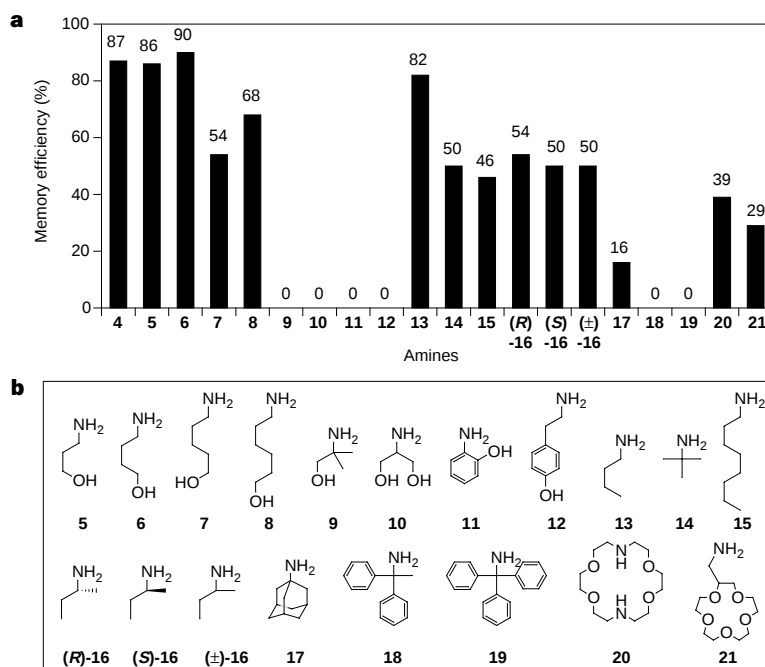


Figure 3 The memory of the macromolecular helicity of poly-**1** with various amines. **a**, Memory efficiencies of the macromolecular helicity of poly-**1** induced by (*R*)-**2** using a series of achiral amino alcohols (**4**–**12**) and amines (**13**–**21**). Memory efficiencies (%) were estimated based on the ICD values at 374 nm, $[\theta]_{374}$, just after the GPC fractionation of a solution of poly-**1**–(*R*)-**2** complex in the

presence of 50 equiv. amino alcohols (**4**–**10**) or in the absence of amines (**11**–**21**) using 0.8 M (**4**–**7**, **10**), 0.4 M (**8**), 0.2 M (**11**) or 0.008 M (**12**–**21**) amines in DMSO as the mobile phase. $[\theta]_{374} = 2.8 \times 10^4$ is used as the base value for the original poly-**1**–(*R*)-**2** complex. **b**, Structures of the amines.

(*R*)-2 (1 equiv.) to a solution of poly-1–4 complex (4; 50 equiv.) did not induce any CD at all. These results strongly indicate that the macromolecular helicity of the poly-1 induced by the (*R*)-2 is retained in the presence of achiral 4. When (*S*)-2 is used instead, poly-1 with the opposite macromolecular helicity is observed when (*S*)-2 is replaced with 4.

Although the memory of the macromolecular helicity of poly-1 assisted by 4 is not perfect just after the GPC fractionation (87%)—based on the original ICD value of the poly-1–(*R*)-2 complex ($[\theta]_{374} = 2.8 \times 10^4$)—it was automatically amplified (repaired) with time in the presence of achiral 4. (Here $[\theta]_{374}$ indicates the ICD intensity at 374 nm.) The CD intensity of the poly-1 increased and reached a plateau value ($[\theta]_{374} = 2.7 \times 10^4$) after 13 days at ambient temperature.

The memory lasted for an extremely long time and hardly changed over 3 months at ambient temperature (20–25 °C); after 3 months $[\theta]_{374}$ showed a decrease of ~5%. Based on the decrease in the CD intensity, the half-life of the memory is roughly estimated to be over 4 years at 20–25 °C. At higher temperatures (50–80 °C), the poly-1 also kept the memory for 300 h at least at 50 °C and 10 h at 80 °C, although at higher temperatures, poly-1 gradually decomposes or isomerizes accompanied by a decrease in the UV–visible intensity. The cooperative hydrogen-bond formation of the hydroxy group of 4 with a carboxy residue of poly-1 as well as the acid–base interaction¹⁷, must facilitate this good long-term memory of the one-handed helical conformation of poly-1. We consider that during the exchange reaction between the bound (*R*)-2 and 4, the latter may come into a chiral binding site (carboxy groups of poly-1) with a chiral *gauche-staggered* conformation so as to keep the helical conformation with the same helix-sense induced by the (*R*)-2. The bound 4 may have a conformation similar to that of the bound (*R*)-3 which can induce the same helix-sense of poly-1.

We further investigated the memory efficiencies of the macromolecular helicity of poly-1 induced by (*R*)-2 using a series of achiral amino alcohols (5–12) and amines (13–21) (Fig. 3). The magnitude of the ICD of the memorized poly-1 is dependent on the number of methylene groups of the amino alcohols (compounds 4–8). In contrast, the helical poly-1 lost its chiral memory on complexation with amino alcohols 9–12. The memory efficiencies, however, are almost independent of the binding affinity to a carboxy group; *K*s were not significantly different from each other ($K(M^{-1}) = 1,522 \pm 90$ (5), 867 ± 113 (6), $1,241 \pm 112$ (7), 470 ± 10 (9) and 774 ± 129 (10)).

As well as the amino alcohols, various primary amines also work as small, achiral chaperoning molecules to assist in the memory of the macromolecular helicity. A simple amine, *n*-butylamine (13), shows a high memory efficiency, but the memory steadily declined at ambient temperature (20–25 °C): the CD intensity showed a decrease of 15% after 51 days, and the half-life may be calculated to be about 200 days, which is much shorter than that of the poly-1–4 system (over 4 years). At 50–80 °C, the memory was lost rapidly and nonlinearity (non-first-order decay) occurred during the initial stage: the initial half-lives are about 30 h, 2 h, 4 min and 1 min at 50, 60, 70 and 80 °C, respectively, probably due to the fast exchange between the free and bound 13 at the high temperatures so that the memory rapidly fades. The amine 13 may be good for the short-term memory of the macromolecular helicity, but for long-term memory, the amino alcohols 4–6 are better. The standard Arrhenius analysis to estimate the thermodynamic parameters is not applicable to this system because of the nonlinear relation between the kinetics and temperature. Similar macromolecular helicity amplification (repair) of the memorized poly-1 was also observed for 5 and 6, but not observed for the other achiral amines. A wide variety of other achiral and chiral amines can be used to construct new supramolecular assemblies with macromolecular helicity—for instance the crown ethers (20, 21 in Fig. 3), which may form a helical array around the poly-1 strand.

We do not yet understand the mechanism of this ‘chaperoning’ of helicity at the molecular level, but the memory effect must be governed kinetically—that is, by the relative rate of exchange of the bound (*R*)-2 with achiral molecules and the atropisomerization (helicity change through bond rotation) process of poly-1. □

Methods

Poly-1 was prepared according to previously reported methods¹⁴; its number-average molecular mass was 1.7×10^4 Da as determined by GPC with polystyrene standards in tetrahydrofuran as its methyl ester. Dry DMSO and dry amines were used in all experiments. CD spectra were measured in a 0.1-, 1.0-, 2.0-, 5.0- or 10.0-mm quartz cell using a J-720 L spectropolarimeter (Jasco, Hachioji, Japan). The concentration of poly-1 was calculated based on the monomer units, and was corrected using the ϵ (molar absorptivity) of poly-1 ($\epsilon_{400} = 2,840 \text{ cm}^{-1} \text{ M}^{-1}$).

GPC fractionation experiments. GPC fractionation was performed using a PU-980 liquid chromatograph (Jasco) equipped with a UV-visible (300 nm; Jasco 970-UV) detector. A KF-806L (Shodex, Tokyo, Japan) GPC column, was connected and 0.8 M (4–7, 9, 10), 0.4 M (8), 0.2 M (11) or 0.008 M (12–21) amines in DMSO as the mobile phase were used at a flow rate of 1.0 ml min^{-1} . The concentration of the amines in the mobile phase may influence the memory efficiency, but we found no significant changes in the memory efficiency in the concentration ranges of 0.8–0.016 M for 4 and 0.8–0.008 M for 13. The addition of excess amines to the poly-1–(*R*)-2 complex before the GPC fractionation is not necessary for the macromolecular helicity memory, but it does affect the memory efficiency, depending on the amines. For instance, the addition of 50 equiv. of 4–7 before the GPC fractionation tended to increase the memory efficiency, while 13 caused a slight decrease in the memory efficiency. Therefore, achiral amino alcohols (50 equiv., 4–10) were added before GPC fractionation, but other achiral amines were not added to the solution of the poly-1–(*R*)-2 complex before GPC fractionation. The use of achiral amines or amino alcohols as the mobile phase component is definitely required for good memory, because the poly-1 fraction separated from the solution of the poly-1–(*R*)-2 complex with 50 equiv. of 4 by GPC using pure DMSO showed a decrease in the ICD intensity (one-third). We note that the poly-1 fraction separated from the poly-1–(*R*)-2 complex using pure DMSO as the mobile phase showed no ICD, indicating that the (*R*)-2 complexed with poly-1 is completely removed during the GPC fractionation.

Received 10 December 1998; accepted 29 March 1999.

- Lehn, J. M. *Supramolecular Chemistry* (VCH, Weinheim, 1995).
- Constable, E. C. Oligopyridines as helical ligands. *Tetrahedron* **46**, 10013–10059 (1992).
- Amabilino, D. B. & Stoddart, J. F. Interlocked and intertwined structures and superstructures. *Chem. Rev.* **95**, 2725–2828 (1995).
- Piguet, C., Bernardinelli, G. & Hopfgartner, G. Helicates as versatile supramolecular complexes. *Chem. Rev.* **97**, 2005–2062 (1997).
- Okamoto, Y. & Nakano, T. Asymmetric polymerization. *Chem. Rev.* **94**, 349–372 (1994).
- Nolte, R. J. M. Helical poly(isocyanides). *Chem. Soc. Rev.* **23**, 11–19 (1994).
- Green, M. M. *et al.* A helical polymer with a cooperative response to chiral information. *Science* **268**, 1860–1866 (1995).
- Rowan, A. E. & Nolte, R. J. M. Helical molecular programming. *Angew. Chem. Int. Edn.* **37**, 703–706 (1998).
- Schlitzer, D. S. & Novak, B. M. Trapped kinetic states, chiral amplification and molecular chaperoning in synthetic polymers: chiral induction in polyguanidines through ion pair interactions. *J. Am. Chem. Soc.* **120**, 2196–2197 (1998).
- Ruokolainen, J. *et al.* Switching supramolecular polymeric materials with multiple length scales. *Science* **280**, 557–560 (1998).
- Tsukruk, V. V. Assembly of supramolecular polymers in ultrathin films. *Prog. Polym. Sci.* **22**, 247–311 (1997).
- Pu, L. The study of chiral conjugated polymers. *Acta Polym.* **48**, 116–141 (1997).
- Okamoto, Y. & Yashima, E. Polysaccharide derivatives for chromatographic separation of enantiomers. *Angew. Chem. Int. Edn.* **37**, 1020–1043 (1998).
- Yashima, E., Matsushima, T. & Okamoto, Y. Chirality assignment of amines and amino alcohols based on circular dichroism induced by helix formation of a stereoregular poly(4-carboxyphenylacetylene) through acid-base complexation. *J. Am. Chem. Soc.* **119**, 6345–6359 (1997).
- Lindsey, J. S. Self-assembly in synthetic routes to molecular devices. Biological principles and chemical perspectives: a review. *New J. Chem.* **15**, 153–180 (1991).
- Harada, N. & Nakanishi, K. *Circular Dichroic Spectroscopy—Exciton Coupling in Organic Stereochemistry* (University Science Books, Mill Valley, CA, 1983).
- Manabe, K., Okamura, K., Date, T. & Koga, K. Effects of ion-pair structure on relative basicity in chloroform: acid-base equilibria controlled by steric repulsion, π -stacking interactions, and hydrogen bonding within an ion pair. *J. Am. Chem. Soc.* **115**, 5324–5325 (1993).

Acknowledgements. We thank Y. Maeda for his preliminary experimental support. We thank T. Sato, K. Akiyoshi and T. Aida for discussions. This work was supported in part by Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture, Japan.

Correspondence and requests for materials should be addressed to E.Y. (e-mail: yashima@apchem.nagoya-u.ac.jp).

Simulation of recent northern winter climate trends by greenhouse-gas forcing

Drew T. Shindell^{*†}, Ron L. Miller^{†‡}, Gavin A. Schmidt^{*†} & Lionel Pandolfo[§]

^{*} NASA Goddard Institute for Space Studies, 2880 Broadway, New York, New York 10025, USA

[†] Center for Climate Systems Research, Columbia University, New York, New York 10025, USA

[‡] Department of Applied Physics, Columbia University, New York, New York 10027, USA

[§] Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, BC, V6T 1Z4 Canada

The temperature of air at the Earth's surface has risen during the past century¹, but the fraction of the warming that can be attributed to anthropogenic greenhouse gases remains controversial. The strongest warming trends have been over Northern Hemisphere land masses during winter, and are closely related to changes in atmospheric circulation. These circulation changes are manifested by a gradual reduction in high-latitude sea-level pressure, and an increase in mid-latitude sea-level pressure associated with one phase of the Arctic Oscillation (a hemisphere-scale version of the North Atlantic Oscillation)². Here we use several different climate-model versions to demonstrate that the observed sea-level-pressure trends, including their magnitude, can be simulated by realistic increases in greenhouse-gas concentrations. Thus, although the warming appears through a naturally occurring mode of atmospheric variability, it may be anthropogenically induced and may continue to rise. The Arctic Oscillation trend is captured only in climate models that include a realistic representation of the stratosphere, while changes in

ozone concentrations are not necessary to simulate the observed climate trends. The proper representation of stratospheric dynamics appears to be important to the attribution of climate change, at least on a broad regional scale.

While the troposphere has been warming in recent decades, the stratosphere has been steadily cooling^{3–5}, the strength of the stratospheric jet has been increasing⁶, and there are indications that the strength of the polar vortex is also increasing⁷. These trends may be related, as observational studies indicate a dynamical coupling between the lower stratosphere and the troposphere^{6–10}. A coupled troposphere/stratosphere mode of internal variability has also been reproduced in climate models^{11,12}.

It is difficult to determine from observations whether these trends, associated with internal modes of variability, are anthropogenically forced^{13,32} or temporary and thus coincidental to the trend in increasing greenhouse-gas concentrations¹⁴. Climate models can help to distinguish between these possibilities. Here we present the response of a number of climate models to an increase in greenhouse-gas concentrations. Each model is based on the NASA Goddard Institute for Space Sciences (GISS) atmospheric general circulation model^{15–17}, and is described in Table 1 and Fig. 1. Briefly, the stratospheric models SC, SG and SO are respectively a control simulation, a simulation with greenhouse-gas forcing, and a simulation with polar ozone and greenhouse gases. The tropospheric models T8G and T4G are greenhouse-gas forcing simulations. Both versions have a low model top compared with the stratospheric models, while T4G has greater horizontal resolution.

Interannual variability of Northern Hemisphere wintertime sea-level pressure (SLP) in both the observations and models is dominated by a nearly axisymmetric spatial pattern centred over the Arctic, as shown by the leading empirical orthogonal function (EOF)¹⁸ in Fig. 2a–c. This observed pattern is defined as the Arctic Oscillation (AO)². The classic North Atlantic Oscillation contrast between mid-latitude and high-latitude SLP is reproduced in the models (Fig. 2b–c) although the observed structure in the North Pacific area is less evident¹⁹.

There is a trend towards decreasing high-latitude SLP and

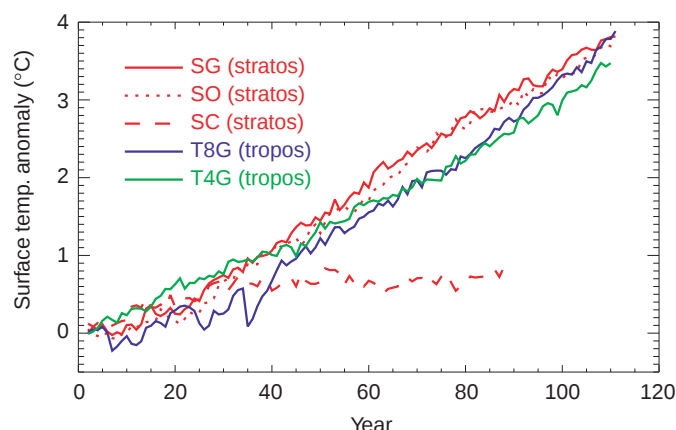


Figure 1 Global annual-average surface temperature anomalies (°C). The model simulations are described in Table 1.

Table 1 Model descriptions and results

Model	Resolution	Model top mbar (km)	Ocean	Length (yr)	Forcing	Slope (mbar per decade)	r^2	Ref.
SC	8° × 10°, 23 levels	0.002 (85)	ML	89	Control	0.03 ± 0.19	0.06	22
SG	8° × 10°, 23 levels	0.002 (85)	ML	111	GHG	0.32 ± 0.10	0.27	22
SO	8° × 10°, 23 levels	0.002 (85)	ML	111	GHG + O ₃	0.31 ± 0.09	0.30	22
T8G	8° × 10°, 9 levels	10 (30)	ML	111	GHG	0.03 ± 0.08	0.01	31
T4G	4° × 5°, 9 levels	10 (30)	OGCM	110	GHG	0.10 ± 0.12	0.03	17

ML is a mixed-layer ocean with diffusion into the deep ocean and fixed ocean heat transport, while OGCM is a fully coupled dynamic ocean. SG and SO were initialized at the end of the control integration SC. All the models were forced with an increasing greenhouse-gas concentration based on observations from 1959 to 1984, after which an increase similar to the IPCC projection IS92a was used²². Aerosol concentrations were constant in all simulations. The slope is the linear least-squares fit to the leading principal component of wintertime Northern Hemisphere SLP (compare Fig. 2) over the entire length of the simulations (except for SC, where the first 20 years of spin-up are discarded). The error estimate corresponds to the 95% confidence level, assuming as a null hypothesis that the detrended time series of successive wintertime values is a first-order auto-regressive process.

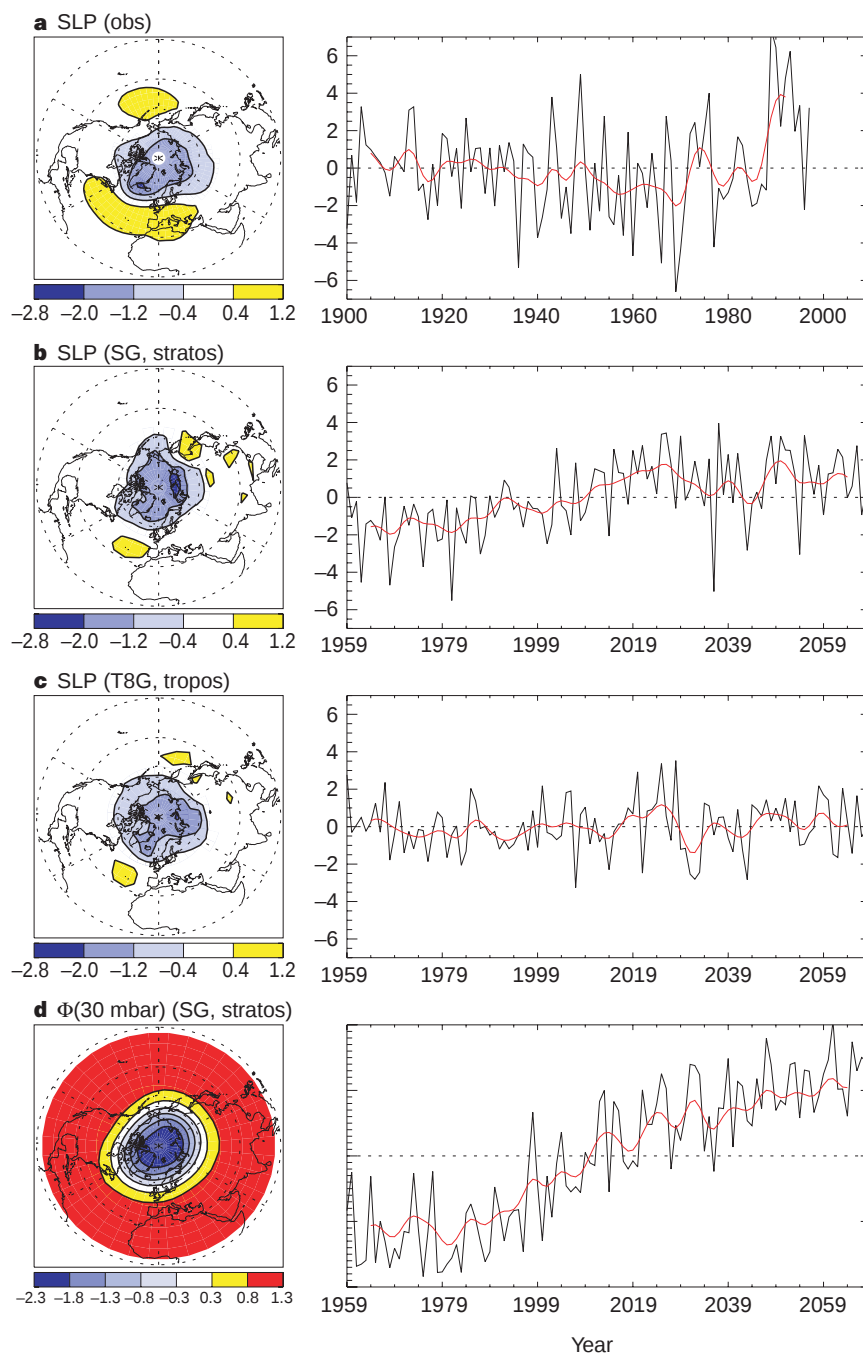


Figure 2 Northern Hemisphere wintertime climate trends. Shown are the leading EOF (left panel) and principal component (PC; right panel) of wintertime (November–April) Northern Hemisphere sea-level pressure (SLP) for **a**, the observations², **b**, the stratospheric model SG (with similar results for SO), **c**, the tropospheric model T8G, and **d**, the leading EOF and PC of wintertime (November–April) Northern Hemisphere 30-mbar geopotential height (m) for the stratospheric model SG. EOFs are a representation of a data set by a number of fixed spatial patterns, each multiplied by a time-varying amplitude (the PC). The patterns are

derived by maximizing the variance corresponding to the pattern multiplied by its PC^{18,30}. In each case, the leading mode is statistically well-separated from succeeding modes whether or not SLP is first detrended. Whereas EOFs and PCs are calculated using the individual months, the PC is plotted as a wintertime average. The PCs are scaled to have units of mbar and the spatial pattern has unit amplitude poleward of 60° N. The red curve represents smoothing of the PC by a 10-year tapered running mean. The first EOF in each case explains 22% (**a**), 13% (**b**), 9% (**c**), and 53% (**d**) of the variance.

increasing mid-latitude SLP in the stratospheric models (SG, SO; Fig. 2b) that mirrors the rise in globally-averaged model temperature (Fig. 1) and which resembles the observed SLP trends of recent decades. Following Thompson and Wallace², we define the time series corresponding to the leading EOF (Fig. 2b) as the ‘index’ of the model’s AO. A linear fit to the AO index has a slope significantly different from zero (Table 1), and the magnitude of the trend (~0.8 mbar per decade over 50 years, starting after 20 years of forcing) resembles the observed value (~1 mbar per decade over the

past 30 years). Interdecadal variability is large, and the AO trend might be overlooked were the simulation not extended beyond the initial few decades. The AO index exhibits little additional increase in the final decades of the simulation, despite the continued increase in greenhouse forcing and surface air temperature.

We find that the observed wintertime SLP trend is reproduced only by models that resolve stratospheric dynamics (Fig. 2b). In contrast, models lacking a dynamical stratosphere (including T8G and T4G) uniformly fail to simulate the observed trend in the AO,

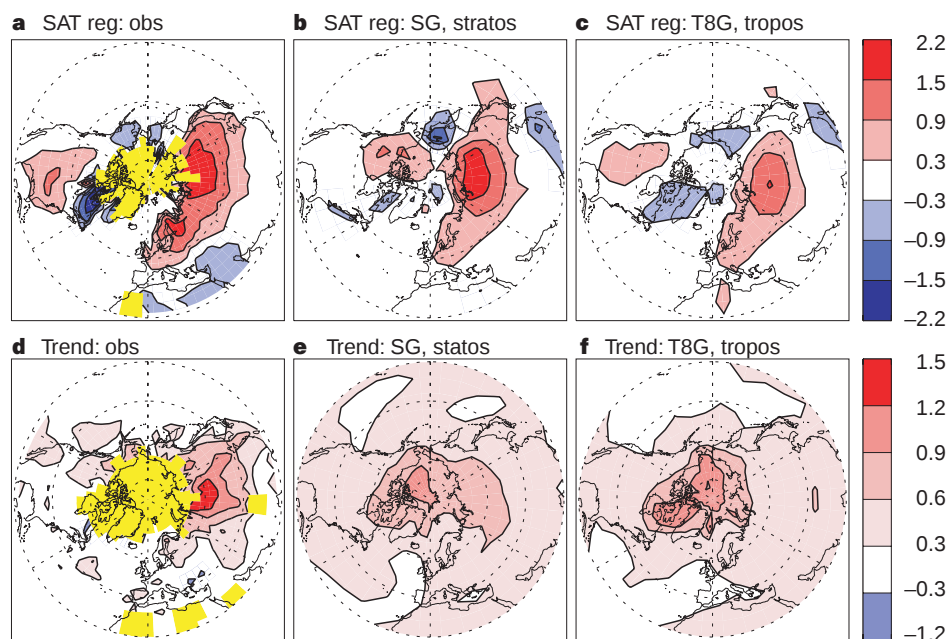


Figure 3 Surface air temperature (SAT) and the Arctic Oscillation (AO). Top row, regression of the AO index upon surface air temperature (°C) for **a**, the observations (ref. 2) (yellow denotes no data), **b**, the stratospheric model SG and **c**, the tropospheric model T8G. Bottom row, linear trend (°C per decade) in

wintertime (November–April) surface air temperature for **d**, the observations, **e**, the stratospheric model SG and **f**, the tropospheric model T8G. The equivalent plots for the stratospheric model SO are very similar to that for SG.

even though this mode is present and dominates the wintertime variability of each model (Fig. 2c). A greater surface warming at high latitudes, the typical response of a general circulation model to increasing greenhouse gases^{11,15–17,20}, does lead to decreased SLP in the Arctic relative to mid-latitudes over time, but these changes are only weakly correlated with the AO. In fact, the trend of the AO time series is not statistically distinct from zero in either tropospheric model (Table 1). Although these models contain two dynamical layers above the tropopause, this is insufficient to resolve stratospheric dynamics. The inability of these tropospheric models to produce an AO trend is not affected by changes in horizontal resolution, physical parametrizations or the ocean component. This behaviour suggests that stratospheric dynamics must be accounted for in the problems of detection and prediction of future warming.

Observed interannual variations in stratospheric circulation are well correlated with the AO (ref. 2), and this relationship is reproduced by the stratospheric models. Figure 2d shows the leading EOF of 30-mbar geopotential height in model SG. The corresponding time variation has a high correlation with the AO index, irrespective of whether each field is first detrended, on both interannual and multi-decadal timescales. A reduction in Arctic SLP is associated with a deepening of the stratospheric polar vortex and a strengthening of the polar night jet.

Polar ozone depletion is a candidate for driving observed stratospheric trends^{2,20,21}, which are correlated with the AO index. However, the simulation SG, where ozone changes are absent, exhibits the same increase in the AO index as does simulation SO, where polar ozone chemistry is calculated through an interactive photochemical parametrization²². To identify the effect of ozone forcing, EOFs of wintertime SLP (for simulation SO) were recalculated for two separate periods: December and January versus March and April. Although ozone forcing is largest during the latter period, the multi-decadal trend in the AO index, evident in Fig. 2b, is present only in the December to January variability. This result, along with the nearly identical AO trends in models SO and SG, suggests that ozone forcing is not necessary to increase the AO index or to strengthen the stratospheric polar vortex.

In addition to its connection to the stratosphere, the AO index is correlated with the recent rise in Northern Hemisphere land temperatures that is cited as evidence of anthropogenic warming^{23,24}. The decrease in high-latitude SLP and increase in mid-latitude SLP is associated with increased wintertime advection of relatively warm oceanic air across northern Asia, causing a local rise in temperature². Similarly, onshore advection warms North America, and the temperature falls in the vicinity of the Bering Strait and the eastern Canadian Arctic owing to the offshore flow of cold continental air.

To examine the relationship between the AO index and surface air temperature, we detrended both fields and regressed the former upon the latter. Detrending is used to remove any spurious correlation that might arise from the trend in both time series. The observed regression pattern (Fig. 3a) is reproduced by all models. Figure 3b and c shows the regression pattern in the stratospheric and tropospheric models, SG and T8G, respectively. A strengthened zonal circulation (corresponding to an increased AO index) is associated with a region of anomalously warm temperature extending from Europe across northern Asia. In addition, cold anomalies are present in the vicinity of the Bering Strait and Labrador. Neither the warm nor the cold anomalies are as extensive as the observed pattern, although better agreement is exhibited by models with higher horizontal resolution (not shown). A comparison of the observed and modelled surface warming is made difficult by the absence of aerosol changes and dynamic ocean feedbacks in these models. In addition, the model trends are estimated over a longer period than the observations, increasing their signal-to-noise ratio. However, Fig. 3d–f suggests that models reproducing the observed AO trend will exhibit enhanced wintertime warming over northern Eurasia as observed.

In the stratospheric models, increasing greenhouse gases have a significant influence on the propagation of planetary waves from the troposphere into the stratosphere^{12,22}. The latitudinal temperature gradient near the tropopause increases as the tropics and mid-latitudes warm, while the high latitudes cool, enhancing the zonal wind at these levels. This results in a deflection of high-latitude wave

energy away from the lower stratosphere. The zonally-averaged vertical circulation is then enhanced, with greater rising motion in the polar latitudes and descent between 40° and 55° N, continuously from the surface to the middle stratosphere, in accord with the deep barotropic nature of the AO (ref. 2). This results in a cooling northward of about 55° N, and a warming at lower latitudes, and a corresponding increase in the westerly zonal wind around 55° N from the surface up to about the 1-mbar level. The net result is a strengthened polar vortex aloft, while the enhanced zonal circulation at the surface leads to a greater advection of warm air onto land, and of cold air over the oceans. Thus stratospheric dynamics modulates the propagation of tropospheric energy, in addition to altering stratospheric energy itself. The influence of the stratosphere on the troposphere is consistent with the observed co-variability between the two levels². Furthermore, a recent analysis shows that, on average, the observed AO signature propagates downwards from the stratosphere reaching the surface after ~3 weeks (M. Baldwin and T. Dunkerton, personal communication).

The GISS stratospheric models are distinguished from their 'tropospheric' counterparts by their ability to generate a realistic climatology in the stratosphere. This is influenced by the higher upper boundary (0.002 mbar versus 10 mbar), greater vertical resolution above the tropopause, and the inclusion of different physics (for example, gravity wave drag), among other factors. Other general circulation models also show a trend in the AO index in response to increasing greenhouse gases (ref. 20 and R. McDonald, personal communication). However, the trends are weaker than those seen in the GISS stratospheric models or in the observations. These general circulation models, like the GISS tropospheric models, have only limited representations of the stratosphere. An upper boundary in the middle stratosphere (~5–10 mbar) limits a model's ability to properly simulate planetary wave propagation^{12,25,26}, which might affect the models' ability both to produce an overall AO trend and to reproduce how the total SLP trend projects upon the AO.

The spatial structure of the Arctic Oscillation is reproduced as the leading wintertime mode of Northern Hemisphere SLP in all the GISS models studied. Furthermore, an increase in the AO index consistent with the recent observed trend occurs in response to anthropogenic greenhouse gases in models containing stratospheric dynamics. This suggests that the observed trend is a signature of anthropogenic greenhouse gases and that trends in surface air temperature associated with the AO should be attributed to human activities, rather than to natural variability. Thus, greenhouse gases may induce climate change through dynamical effects, in addition to direct radiative forcing. Despite the absence of ozone forcing in one stratospheric model, both models exhibit strengthening of the stratospheric polar vortex in association with the decrease in Arctic SLP, suggesting that ozone is not fundamental to the observed trend. The failure of tropospheric models to reproduce the observed AO trend suggests that stratospheric dynamics are crucial to certain important aspects of tropospheric variability, at least in the models analysed here. As changes in SLP are associated with anomalies in surface air temperature, the inclusion of stratospheric dynamics in a model may be fundamental to the simulation of present warming, along with the prediction of future climate change. Although hemispherically averaged warming associated with the AO trend is a small component of the total, it has a distinct regional signal which may be important for detection of the anthropogenic influence on surface air temperatures. There may be many ways to excite the AO. Detection and attribution of the effect of a change in the AO will probably depend on proper simulation of the dominant mechanisms governing its behaviour.

We suggest that the stratosphere influences climate at the surface by modulating energy propagation out of the troposphere. This mechanism implies that changes to the stratosphere, whether induced by variations in greenhouse gases, solar flux, ozone, or

volcanic forcing^{21,27,28}, may all affect surface climate. It remains unclear whether, and through what mechanism, the AO trend saturates towards the end of the simulation (Fig. 2b), despite the continued increase in forcing and surface air temperature. Perhaps the stratospheric polar vortex has strengthened and stabilized to such an extent that virtually all planetary waves are refracted away. The relative importance of the various factors influencing the quality of the model's stratospheric simulation (for example, the height of the model's upper boundary, the vertical resolution above the tropopause) will become clearer with further study. An Antarctic counterpart to the AO trend is also exhibited by both stratospheric models, with amplitude comparable to Arctic values. Such a trend has recently been identified in the observations²⁹. □

Received 23 October 1998; accepted 29 March 1999.

- Hansen, J., Ruedy, R., Sato, M. & Reynolds, R. Global surface air temperature in 1995: Return to pre-Pinatubo level. *Geophys. Res. Lett.* **23**, 1665–1668 (1996).
- Thompson, D. W. J. & Wallace, J. M. The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.* **25**, 1297–1300 (1998).
- Oort, A. H. & Liu, H.-Z. Upper-air temperature trends over the globe. *J. Clim.* **6**, 292–307 (1993).
- Spencer, R. W. & Christy, J. R. Precision lower stratospheric temperature monitoring with the MSU: Technique, validation, and results 1979–1991. *J. Clim.* **6**, 1194–1204 (1993).
- Pawson, S., Labitzke, K. & Leder, S. Stepwise changes in stratospheric temperature. *Geophys. Res. Lett.* **25**, 2157–2160 (1998).
- Kodera, K. & Koide, H. Spatial and seasonal characteristics of recent decadal trends in the Northern Hemisphere troposphere and stratosphere. *J. Geophys. Res.* **102**, 19433–19447 (1997).
- Zurek, R. W., Manney, G. L., Miller, A. J., Gelman, M. E. & Nagatani, R. M. Interannual variability of the north polar vortex in the lower stratosphere during the UARS mission. *Geophys. Res. Lett.* **23**, 289–292 (1996).
- Baldwin, M. P., Cheng, X. & Dunkerton, T. J. Observed correlations between winter mean tropospheric and stratospheric circulation anomalies. *Geophys. Res. Lett.* **21**, 1141–1144 (1994).
- Spencer, J. & Graf, H.-F. The statistical connection between tropospheric and stratospheric circulation of the Northern Hemisphere in winter. *J. Clim.* **8**, 2281–2295 (1995).
- Kodera, K., Chiba, M., Koide, H., Kotoh, A. & Nikaidou, Y. Interannual variability of the winter stratosphere and troposphere in the Northern Hemisphere. *J. Meteorol. Soc. Jpn* **74**, 365–382 (1996).
- Kitoh, A., Koide, H., Kodera, K., Yukimoto, S. & Noda, A. Interannual variability in the stratosphere-troposphere circulation in a coupled ocean-atmosphere GCM. *Geophys. Res. Lett.* **23**, 543–546 (1996).
- Rind, D., Shindell, D., Lonerger, P. & Balachandran, N. K. Climate change and the middle atmosphere. Part III: The doubled CO₂ climate revisited. *J. Clim.* **11**, 876–894 (1998).
- Palmer, T. N. A nonlinear dynamical perspective on climate change. *Weather* **48**, 314–326 (1993).
- Hurrell, J. W. Influence of variations in extratropical wintertime teleconnections on Northern Hemisphere temperature. *Geophys. Res. Lett.* **23**, 665–668 (1996).
- Hansen, J. *et al.* Efficient three-dimensional global models for climate studies: Models I and II. *Mon. Weath. Rev.* **111**, 609–662 (1983).
- Rind, D., Suozzo, R., Balachandran, N. K., Laci, A. & Russell, G. The GISS global climate/middle atmosphere model. Part I: Model structure and climatology. *J. Atmos. Sci.* **45**, 329–370 (1988).
- Russell, G. L., Miller, J. R. & Rind, D. A coupled atmosphere-ocean model for transient climate change. *Atmosphere-Ocean* **33**, 683–730 (1995).
- Bretherton, C. S., Smith, C. & Wallace, J. M. An intercomparison of methods for finding coupled patterns in climate data. *J. Clim.* **5**, 541–560 (1992).
- Wallace, J. M. & Gutzler, D. S. Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Weath. Rev.* **109**, 784–812 (1981).
- Graf, H.-F., Kirchner, I. & Perlwitz, J. Changing lower stratospheric circulation: The role of ozone and greenhouse gases. *J. Geophys. Res.* **103**, 11251–11261 (1998).
- Volodin, E. M. & Galin, V. Interpretation of winter warming at Northern Hemisphere continents 1977–94. *J. Clim.* (in the press).
- Shindell, D. T., Rind, D. & Lonerger, P. Increased polar stratospheric ozone losses and delayed eventual recovery due to increasing greenhouse gas concentrations. *Nature* **392**, 589–592 (1998).
- Wallace, J. M., Zhang, Y. & Bajuk, L. Interpretation of interdecadal trends in Northern Hemisphere surface air temperature. *J. Clim.* **9**, 249–259 (1996).
- Broccoli, A. J., Lau, N. C. & Nath, M. J. The cold ocean warm land pattern: Model simulation and relevance to climate change detection. *J. Clim.* **11**, 2743–2763 (1998).
- Boville, B. A. & Cheng, X. Upper boundary effects in a general circulation model. *J. Atmos. Sci.* **45**, 2591–2606 (1988).
- Austin, J., Butchart, N. & Swinbank, R. S. Sensitivity of ozone and temperature to vertical resolution in a GCM with coupled stratospheric chemistry. *Q.J.R. Meteorol. Soc.* **123**, 1405–1431 (1997).
- Shindell, D. T., Rind, D., Balachandran, N., Lean, J. & Lonerger, P. Solar cycle variability, ozone, and climate. *Science* **284**, 305–308 (1999).
- Kodera, K. On the origin and nature of interannual variability of the winter stratospheric circulation in the Northern Hemisphere. *J. Geophys. Res.* **100**, 14077–14087 (1995).
- Gong, D. & Wang, S. Definition of Antarctic oscillation index. *Geophys. Res. Lett.* **26**, 459–462 (1999).
- North, G. R., Bell, T. L., Cahalan, R. F. & Moeng, F. J. Sampling errors in the estimation of empirical orthogonal functions. *Mon. Weath. Rev.* **110**, 699–706 (1982).
- Rosenzweig, C. *et al.* Building Linkages among Climate, Impacts, and Economics: A Multi-dimensional Approach to Integrated Assessment (Columbia Earth Institute & NASA GISS, New York, NY, 1998).
- Corti, S., Molteni, F. & Palmer, T. N. Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature* **398**, 799–802 (1999).

Acknowledgements. We thank J. Hansen, D. Rind and G. Russell for making available their numerical experiments, D. Thompson for providing analyses of the observations and P. Lonerger for assistance with model output. We also thank P. Stone and J. M. Wallace for their comments on earlier drafts of this Letter. G.A.S. was supported by a NOAA postdoctoral fellowship in Climate and Global Change administered by the UCAR Visiting Scientist Program. R.L.M. is supported by the NOAA Atlantic Climate Change Program. Stratospheric modeling at GISS is supported by the NASA Atmospheric Chemistry Modeling and Analysis Program.

Correspondence and requests for materials should be addressed to D.T.S. (e-mail: dshindell@giss.nasa.gov).

Diamonds in volcanoclastic komatiite from French Guiana

Ramon Capdevila*, Nicholas Arndt*‡, Jacques Letendre†‡ & Jean-François Sauvage

* Géosciences, UPR 4661 CNRS, Université de Rennes 1, 35042 Rennes cedex, France

† Golden Star Resources Ltd, 57 High Street, Kingston, PO Box 10313, Georgetown, Guyana

§ Guyanor Ressources SA, BP 750, 5, Lotissement Calimbé II, Route du Tigre, 97337 Cayenne, French Guiana

‡ Present addresses: Institut Dolomieu, 15 rue M. Gignoux, 38031 Grenoble Cedex, France (N.A.); 6295 Boulevard Lassalle, Verdun, Québec H4H 1P8, Canada (J.L.).

The world's main sources of non-alluvial diamonds are found in ultrapotassic kimberlite¹ or lamproite² diatremes (pipes filled with explosive volcanic debris), most of which have Phanerozoic ages and are located in stable Precambrian cratons. Diamond exploration has therefore tended to focus on such deposits. Microdiamonds are known to occur in metamorphic rocks such as gneiss³ and eclogite⁴ that have equilibrated deep in the mantle and were then tectonically transported to the surface, but such deposits are thought to have little commercial potential. Here we report a new type of diamond occurrence from the Dachine region

in French Guiana for which the host rock is volcanoclastic komatiite—an unusual type of volcanic rock whose composition and origin are quite unlike those of kimberlite and lamproite. These komatiites form part of a Proterozoic island-arc sequence, a tectonic setting distinct from that of all other currently exploited diamond deposits. The discovery of diamonds in volcanoclastic komatiite has implications not only for diamond exploration, but also provides strong evidence that these komatiite magmas originated at depths of 250 km or greater within the Earth.

The Dachine deposit is located in the 2.11 ± 0.09 Gyr Inini greenstone belt⁵ of the Guiana shield (Fig. 1). Following an initial discovery of alluvial diamonds by the French geological survey (BRGM)⁶, geologists of Guyanor Ressources SA demonstrated that the ultramafic host of the deposit is at least 5 km long and 350–1,100 m wide. Bulk samples from this rock contain from <1 to 77 diamonds per kg. Although the diamond population is dominated by microdiamonds, larger diamonds (>1 mm) are locally abundant, their grade reaching 4 carats per m³ in poorly sorted alluvium overlying mineralized bedrock. The largest recovered diamond is ~4.6 mm in diameter.

The primary morphology of diamonds (particularly the presence of cubo-octahedra) and their low carbon-isotope ratios ($\delta^{13}\text{C}$ predominantly -23‰ to -27‰) provide evidence of eclogitic sources⁷. Indicator minerals are also unusual. Although the garnet population is dominated by ilherzolitic types with subordinate subcalcic harzburgitic (G10) and eclogitic groups, other minerals commonly associated with kimberlite (Mg-ilmenite, chromian diopside and perovskite) are absent⁸. Chromite cores are poorer in Ti and richer in Mn than those diagnostic of kimberlite and

Table 1 Analyses of diamondiferous volcanoclastic komatiite

Drill hole	1	3	5	5	5	5	6	6	6
Depth (m)	32	29*	25	59	100	138*	41	114*	156*
Composition (wt %)									
SiO ₂	41.1	46.5	38.3	42.7	43.5	38.5	48.2	46.1	46.9
TiO ₂	0.60	0.80	0.81	0.72	0.76	0.60	0.62	0.59	0.56
Al ₂ O ₃	6.44	8.75	8.34	7.07	7.89	6.61	7.24	4.82	4.35
Fe ₂ O ₃ †	11.0	11.6	11.9	11.5	11.9	10.3	11.2	10.7	10.0
MnO	0.15	0.11	0.09	0.13	0.10	0.21	0.10	0.10	0.11
MgO	20.9	17.7	21.8	22.2	20.7	19.1	19.3	23.4	23.0
CaO	6.17	4.53	5.08	4.04	3.49	7.47	3.18	3.93	5.29
Na ₂ O	0.80	1.96	<0.05	<0.05	1.36	0.79	1.91	0.21	0.32
K ₂ O	<0.05	0.72	<0.05	<0.05	<0.05	3.30	<0.05	0.89	0.36
P ₂ O ₅	0.25	0.27	0.19	0.19	0.28	0.21	0.22	0.24	0.19
LOI‡	12.64	7.02	13.20	11.53	9.73	12.38	7.04	8.30	8.20
Total	100.0	99.9	99.6	100.0	99.7	99.5	99.1	99.3	99.4
CO ₂	8.13	2.62	7.44	5.84	4.69	9.49	2.90	2.69	2.64
Elements (p.p.m.)									
Cs	0.176	2.56	0.191	0.481	0.398	8.46	0.405	5.39	3.62
Rb	0.9	25.3	0.7	1.5	1.1	122	0.8	36.0	16.7
Ba	13	181	4.4	11	15	727	11	225	130
Sr	198	137	252	238	245	491	185	184	196
Th	0.36	0.4	1.07	0.84	0.93	0.23	0.66	0.30	0.29
Nb	2.21	2.98	6.87	5.79	6.74	2.05	3.57	2.36	2.44
Zr	41	55	58	54	55	41	50	34	38
Y	11.0	13.9	12.8	11.2	13.6	9.48	12.1	8.9	8.6
Cr	1,570	1,675	1,921	1,785	1,813	2,142	1,753	2,088	2,032
Ni	840	631	847	814	687	908	767	913	1,219
La	2.60	4.15	12.49	9.97	12.18	3.99	7.45	3.11	3.60
Ce	7.23	11.09	28.42	23.85	29.13	9.62	16.78	8.04	9.16
Nd	5.70	8.74	15.48	13.09	14.61	6.31	9.94	5.11	5.72
Sm	1.62	2.50	3.11	2.70	3.41	1.77	2.22	1.62	1.33
Eu	0.49	0.67	0.78	0.68	0.72	0.72	0.65	0.51	0.38
Gd	1.53	2.66	2.44	2.25	2.58	1.66	2.30	1.55	1.40
Dy	1.66	2.52	2.33	1.98	2.34	1.69	1.94	1.60	1.36
Er	0.97	1.45	1.23	1.10	1.24	0.89	1.19	0.82	0.84
Yb	1.03	1.38	1.20	1.01	1.20	0.95	1.10	0.80	0.90
Lu	0.15	0.20	0.19	0.18	0.17	0.15	0.18	0.13	0.15

These representative analyses are of samples from Dachine (French Guiana). All analysed samples contain diamonds. Analyses are normalized to 100% on volatile-free basis.

* Metasomatized samples with phlogopite.

† Total iron as Fe₂O₃.

‡ Loss on ignition.

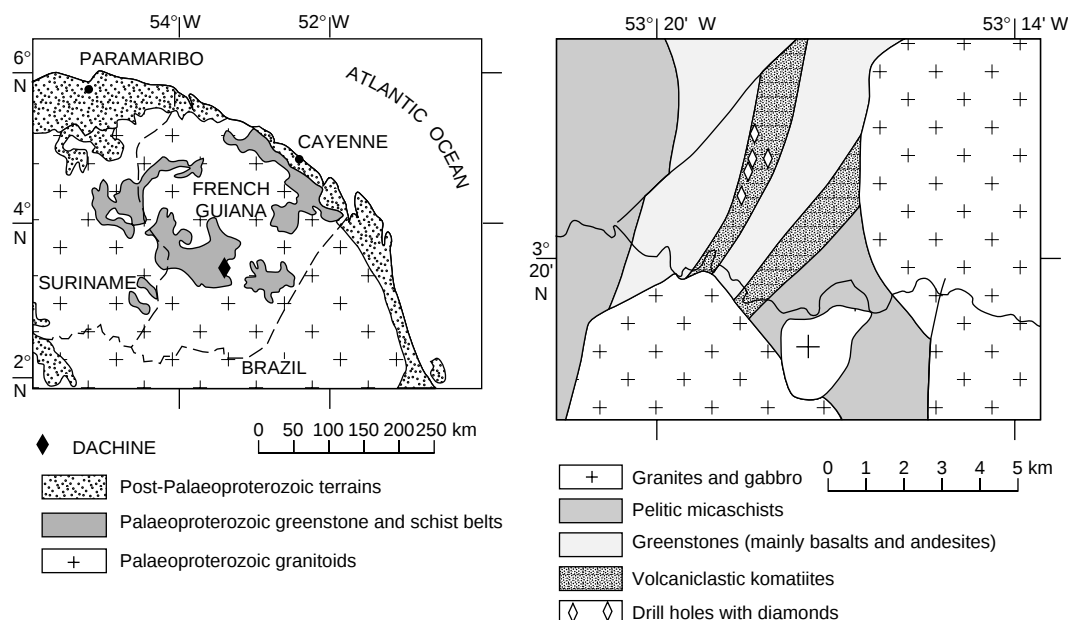


Figure 1 Geological maps. Shown are the location of the Dachine region and the local geology (left), and the distribution of diamond-bearing ultramafic rocks (right).

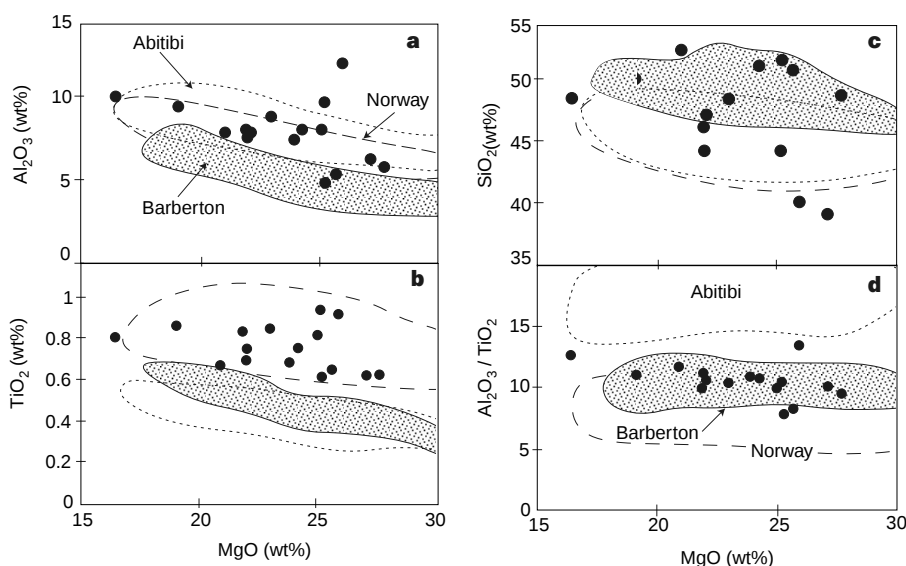


Figure 2 Major-element compositions of Dachine volcaniclastic komatiites. Also shown for comparison are the compositions of komatiites and related rocks from Abitibi²⁰, Barberton²⁰ and Norway¹². Panels **a** and **c** show that the ranges of MgO,

Al_2O_3 and SiO_2 contents are similar in all groups of rocks. Panels **b** and **d** show that in terms of TiO_2 and $\text{Al}_2\text{O}_3/\text{TiO}_2$ the Dachine komatiites most closely match the Barberton and Norwegian examples.

lamproite and are more like spinels from greenstone belts⁹. Low-relief secondary surface features of the diamonds suggest transport in strongly reactive magma with a high $\text{H}_2\text{O}/\text{CO}_2$ ratio⁷. Preliminary Fourier-transform infrared spectrometric results indicate that ~30% of diamonds contain poorly aggregated nitrogen defects; this constrains the time they could have survived at normal mantle temperatures of 1,000–1,400 °C to <200,000 years (I. Chinn, personal communication).

The Inini greenstone belt is dominated by calc-alkaline andesite to rhyolite, and immature sedimentary rocks, all intruded by granitoids including tonalites and trondhjemites. These features indicate an island-arc setting^{5,10}. The ultramafic rocks that host the Dachine diamonds form part of the volcanic package. Most have

been converted to finely foliated albite–carbonate–chlorite–talc schists, but primary volcanic textures are well preserved in some outcrops. The rocks contain ellipsoidal, generally monolithic, ultramafic fragments averaging 1–3 cm (maximum 20 cm) in a fine-grained groundmass. Relict olivine phenocrysts in many fragments eliminate the possibility that the ultramafic composition is due to Mg uptake during metamorphism. The origin of the rocks was either pyroclastic or perhaps hyaloclastic.

The greenstone belt went through two periods of deformation, and was metamorphosed to greenschist facies during trans-Amazonian orogeny. Hydrothermal alteration associated with intrusion of late-tectonic granitoids and gabbros produced large phlogopite porphyroblasts, especially adjacent to gabbroic intrusions.

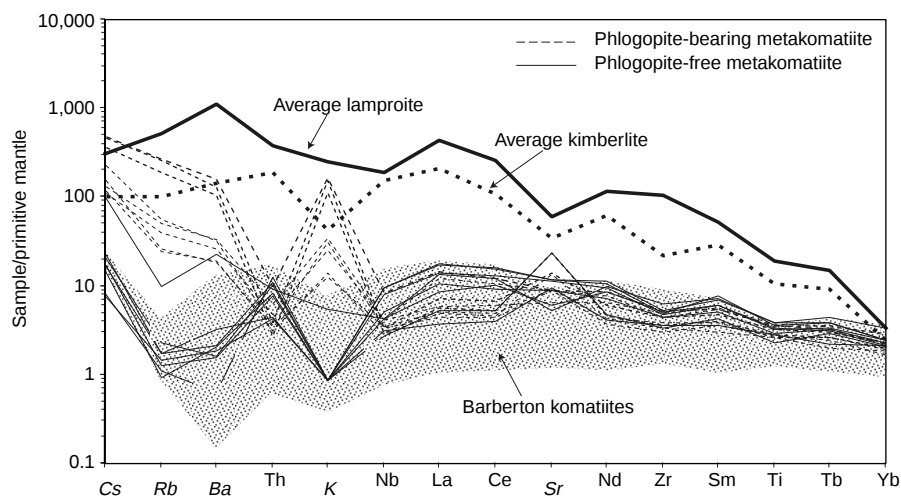


Figure 3 Trace-element compositions of Dachine komatiites, normalized to primitive mantle²¹. Also shown for comparison are patterns for average lamproite and average kimberlite¹², and the compositions of komatiites from Barberton²⁰.

The ultramafic fragmental rocks contain between 17% and 28% MgO (Table 1) and have compositions like komatiites from other regions (Figs 2, 3). Elements such as Al, Ti and Ni, rare-earth elements (REE), and high-field-strength elements (HFSE), which are relatively insoluble in aqueous fluids and immobile during metamorphism, correlate well with one another. More soluble elements such as the alkali and alkaline earth elements, and Eu among the REE, scatter widely (Fig. 3).

Concentrations of immobile elements are very low, similar to those in komatiites, and distinct from kimberlite and lamproite (Fig. 3). Although the high K₂O contents of some Dachine rocks might be taken to indicate a kimberlitic character, Fig. 3 shows that potassium concentrations, like those of Cs, Ba and Rb, vary widely, in contrast to the more coherent behaviour of the immobile elements. The restriction of high concentrations to one set of samples—those with phlogopite porphyroblasts collected near gabbroic intrusions—demonstrates that this characteristic is secondary. It is unlikely that coherent patterns of geochemically different elements (Th, REE, HFSE; Fig. 3) resulted from partial leaching of these elements from originally more enriched ultrapotassic rocks such as kimberlite, and we are confident that abundances of immobile elements are close to those of the original magmas. The low concentrations of compatible elements (Al, Ti and heavy REE) in the Dachine ultramafic volcanoclastic rocks resemble those of Al-depleted or Barberton-type komatiites¹¹, but the closest match is with other occurrences of Precambrian fragmental komatiite^{12–14}.

How do we explain the presence of diamond in komatiite? Ultramafic magma may form through melting of a hot and deep mantle source¹⁵, or a cooler, hydrous, shallower source¹⁶. Although the volcanoclastic nature of the Dachine komatiite might favour the latter explanation, the presence of diamonds argues against it. The Dachine diamonds could not have survived melting of even a relatively cool hydrous mantle source. They probably are xenocrysts, sampled at depths >150 km where diamond is stable, then transported rapidly to the surface. The Al-depleted character of the komatiite indicates that garnet was residual during partial melting of its source, which points to formation at depths >250 km (ref. 17). We propose that primary, anhydrous komatiite magma formed by deep melting, then penetrated hydrated lithosphere beneath the ancient Dachine island arc where it picked up both water and diamonds. As komatiite magma interacted with the relatively cool hydrated base of the mantle wedge, it become hydrous, its tem-

perature and density decreased dramatically and it was propelled to the surface, bringing with it xenocrystic diamonds.

The discovery of diamonds in volcanoclastic komatiite has two main implications. First, it provides a possible explanation for enigmatic occurrences of diamonds throughout the Guianese and West African cratons¹⁸. The source of these diamonds is unknown but some¹⁹ may occur in ultramafic schists like those of the Dachine region. Second, it places significant constraints on the origin of komatiite magmas and the manner in which they interact with hydrated mantle in subduction zones. □

Received 4 December 1998; accepted 29 March 1999.

- Mitchell, R. H. *Kimberlites* (Plenum, New York, 1986).
- Mitchell, R. H. & Bergman, S. C. *Petrology of Lamproites* (Plenum, New York, 1991).
- Sobolev, N. V. & Shatsky, V. S. Diamond inclusions in garnets from metamorphic rocks: a new environment for diamond formation. *Nature* **343**, 742–746 (1990).
- Xu, S. *et al.* Diamond from the Dabie Shan metamorphic rocks and its implication for tectonic setting. *Science* **256**, 80–82 (1992).
- Gruau, G., Martin, H., Leveque, B., Capdevila, R. & Marot, A. Rb–Sr and Sm–Nd geochronology of Lower Proterozoic granite-greenstone terrains in French Guiana, South America. *Precamb. Res.* **30**, 63–80 (1985).
- Picot, P. *Inventaire du Département de la Guyane. Indice Diamantifère de IT 33 (Haut Inini)* 1–45 (BRGM Rep. 82GUY004, Bureau de Recherches Géologiques et Minières, Orléans, France, 1982).
- Letendre, J., McCandless, T. E. & Eastoe, C. J. in *Proc. 7th Int. Kimberlite Conf., Univ. Cape Town, South Africa* 500–502 (1998).
- Bailey, L. M., Helmstaedt, H. H., Peterson, R. C., Mandarin, J. A. & Letendre, J. in *Proc. 7th Int. Kimberlite Conf., Univ. Cape Town, South Africa* 37–39 (1998).
- Griffin, W. L., Fisher, N. I., Friedman, J. H. & Ryan, C. G. Statistical techniques for the classification of chromites in diamond exploration samples. *J. Geochem. Explor.* **59**, 233–249 (1997).
- Vanderhaeghe, O. *et al.* Contrasting mechanism of crustal growth. Geodynamic evolution of the Paleoproterozoic granite-greenstone belts of French Guiana. *Precamb. Res.* **92**, 165–193 (1998).
- Sun, S. S. & Nesbitt, R. W. Petrogenesis of Archean ultrabasic and basic volcanics: evidence from rare earth elements. *Contrib. Mineral. Petrol.* **65**, 301–325 (1978).
- Barnes, S. J. & O'Brien, M. Ti-rich komatiites from northern Norway. *Contrib. Mineral. Petrol.* **105**, 42–54 (1990).
- Saverikko, M. The pyroclastic komatiite complex at Sattasvaara in northern Finland. *Bull. Geol. Soc. Finland* **57**, 55–87 (1985).
- Schaefer, S. J. & Morton, P. Two komatiitic pyroclastic units, Superior Province, northwestern Ontario: their geology, petrography, and correlation. *Can. J. Earth. Sci.* **28**, 1455–1470 (1991).
- Herzberg, C. Depth and degree of melting of komatiite. *J. Geophys. Res.* **97**, 4521–4540 (1992).
- Parman, S., Dann, J., Grove, T. L. & de Wit, M. J. Emplacement conditions of komatiite magmas from the 3.49 Ga Komati Formation, Barberton Greenstone Belt, South Africa. *Earth Planet. Sci. Lett.* **150**, 303–323 (1997).
- Nisbet, E. G., Cheadle, M. J., Arndt, N. T. & Bickle, M. J. Constraining the potential temperature of the Archean mantle: a review of the evidence from komatiites. *Lithos* **30**, 291–307 (1993).
- Janse, A. J. A. & Sheahan, P. A. Catalogue of world wide diamond and kimberlite occurrences: a selective and annotative approach. *J. Geochem. Explor.* **53**, 73–111 (1995).
- Junner, N. R. The diamond deposits of the Gold Coast. *Gold Coast Geol. Surv. Bull.* **12**, 1–52 (1943).
- Lahaye, Y. *et al.* The influence of alteration on the trace-element and Nd isotope compositions of komatiites. *Chem. Geol.* **126**, 43–64 (1995).
- Sun, S.-S. & McDonough, W. F. in *Magnetism in the Ocean Basins* (eds Saunders, A. D. & Norry, M. J.) 313–345 (Spec. Publ. 42, Geol. Soc. London, 1989).

Correspondence and requests for materials should be addressed to N.A. (e-mail: arndt@ujf-grenoble.fr).

Population density affects sex ratio variation in red deer

Loeske E. B. Kruuk^{*†}, Tim H. Clutton-Brock^{*}, Steve D. Albon[‡], Josephine M. Pemberton[†] & Fiona E. Guinness^{*}

^{*} Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK

[†] Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK

[‡] Institute of Terrestrial Ecology, Hill of Brathens, Banchory, Kincardineshire AB31 4BY, UK

Many mammal populations show significant deviations from an equal sex ratio at birth, but these effects are notoriously inconsistent¹. This may be because more than one mechanism affects the sex ratio and the action of these mechanisms depends on environmental conditions. Here we show that the adaptive relationship between maternal dominance and offspring sex ratio previously demonstrated in red deer (*Cervus elaphus*)^{2,3}, where dominant females produced more males, disappeared at high population density. The proportion of males born each year declined with increasing population density and with winter rainfall, both of which are environmental variables associated with nutritional stress during pregnancy. These changes in the sex ratio corresponded to reductions in fecundity, suggesting that they were caused by differential fetal loss. In contrast, the earlier association with maternal dominance is presumed to have been generated pre-implantation. The effects of one source of variation superseded the other within about two generations. Comparison with other ungulate studies indicates that positive associations between maternal quality and the proportion of male offspring born have only been documented in populations below carrying capacity.

It has been argued that, if maternal condition affects the breeding success of male offspring more than that of female offspring, mothers in superior condition should produce more males, whereas those in poorer condition should produce more females⁴. Despite the intuitive appeal of this argument, few studies have provided unequivocal proof of such manipulation, and sex ratio trends in mammalian populations are disconcertingly variable^{1,5}. For example, in ungulate populations, some studies have shown that females in good condition produce significantly greater proportions of sons^{6–11}, whereas others found either no association between condition and offspring sex ratio^{12–14} or a negative association^{15,16}. One plausible explanation for the variation is that the mechanisms generating sex ratio variation are affected by environmental conditions.

Previous studies of the unmanaged red deer population on the Isle of Rum, Scotland, showed that dominant females were

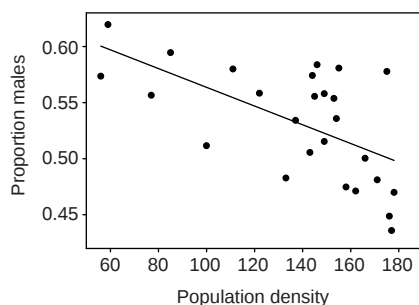


Figure 1 Proportion of males born each year in relation to population density, after correcting for effect of winter rainfall. Regression coefficient: $-0.080 (\pm 0.026 \text{ s.e.})$, $t = -3.12$, d.f. = 23, $P = 0.005$; $r = -0.552$.

consistently more likely than subordinates to produce sons^{2,3}. As maternal dominance affected the breeding success of male offspring more than that of females, this variation in offspring sex ratio was likely to increase a female's expected number of grandchildren⁴. However, rising population density in the study area has resulted in a significant decline in the proportion of males born each year (hereafter, the annual birth sex ratio)¹⁷, which has fallen by approximately 10% since the cessation of culling in 1973 (Fig. 1). This overall decline reflects a decline in the proportion of males born to dominant females (Fig. 2). Fitting individual calf sex as the response variable in a generalized linear mixed model (GLMM; see Methods) showed that there was a significant interaction between population density at the time, and maternal dominance rank: offspring sex ratio was associated with dominance at low but not at high density (Table 1; Fig. 2, inset). Furthermore, a positive association between a female's condition, assessed by kidney fat weight, and the probability of her carrying a male calf has been shown in adjoining areas of the same island, where culling has maintained density at levels equivalent to that at the start of this study¹¹. Thus, the changes in our study population were not associated with a larger-scale temporal trend.

The sex ratio at birth was also correlated with the amount of rainfall between November and January (approximately the second to fourth months of gestation). On average, an extra 100 mm of winter rainfall reduced the annual birth sex ratio in the following spring by 1.3% (Fig. 3). In the GLMM of calf sex, winter rainfall was the only weather variable to have any significant explanatory power (Table 1; see Methods for details of other weather variables considered). If the GLMM was repeated separately for mothers suckling a calf from the previous year (and thus known to be in poorer condition¹⁸) and for those without a calf, winter rainfall was only significant in the former ($\chi^2_{(1)} = 7.5$, $P = 0.006$) and not in the latter ($\chi^2_{(1)} = 0.50$, $P = 0.480$).

As conception occurs in October, the correlation between November–January rainfall and annual birth sex ratio was presumably generated by post-implantation fetal mortality, occurring no earlier than the second month of gestation. This was confirmed by an analysis of female fecundity: the probability that a female calved in a given spring decreased with rainfall in the preceding

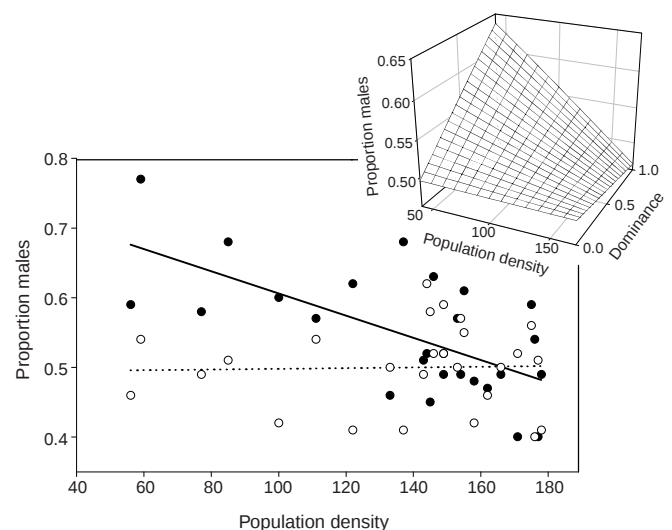


Figure 2 Proportion of males born to dominant versus subordinate females in each year. Dominant females (filled circles and full line) are those with above-median dominance scores; subordinates (open circles and dotted line) are those with below-median scores. For each year, the average of the two values does not necessarily correspond to the population mean in Fig. 1, as dominance rank was not known for all females. Inset: Relation between sex ratio, dominance rank and population density as predicted by the generalized linear mixed model.

Table 1 Generalized linear mixed model of calf sex

Calf sex			
Term	Wald statistic (χ^2)	d.f.	P-value
Mother's age and age ²	0.1	2	0.95
Previous calf	0.3	2	0.86
Mother's dominance	0.1	1	0.75
Winter rainfall	5.5	1	0.019
Population density	3.4	1	0.065
Dominance-density	4.5	1	0.034
Minimal model	Average effect	s.e.	
Mother's dominance	0.051	0.17	
Winter rainfall	-0.00061	0.00026	
Population density	-0.0026	0.0017	
Dominance-density	-0.012	0.0057	

GLMM of calf sex gives the probability that a calf is male; $n = 1,587$. Population density and winter rainfall are for the winter preceding a calf's birth. Wald statistics are from the full model when each term was fitted last in the model; average effects are from the minimal model comprising only the significant terms. Interactions are denoted by a dot between terms. d.f., Degrees of freedom; s.e., standard error.

winter (Table 2). Female fecundity also decreased with rising population density. Although it is impossible to separate the pre- and post-conception effects of population density, we suggest that the nutritional stress of high density, mediated through competitive interactions, is more likely to affect fetal mortality during the winter than female conception rates in the autumn, and thus that the effects of density were also generated by differential fetal mortality. Fecundity increased significantly with dominance ranking ($P = 0.046$) (see also ref. 19). However, unlike the association between dominance and offspring sex ratio, this relationship was not affected by density (Table 2): dominant females were consistently more likely to calve in a given year than subordinates, but at low densities the sex ratio of their offspring differed. This supports previous suggestions that any relationship between dominance and offspring sex ratio was not generated by differential fetal mortality, but instead involved mechanisms acting before implantation: corpus luteum function is affected by social status in red deer²⁰, and the observed difference in fecundity would have to more than double to account for the difference in sex ratios between dominants and subordinates².

It is unclear whether the decline in sex ratio under poor winter conditions represents an adaptive maternal strategy. Both high population density and high November–January rainfall increase juvenile mortality in the population during the winter^{21,22}, an indication that they are associated with nutritional stress. However, juvenile winter mortality occurs mainly in March and April¹⁸, indicating that the effects of the stress are not realized until slightly later, at a time corresponding to the later stages of gestation. It is

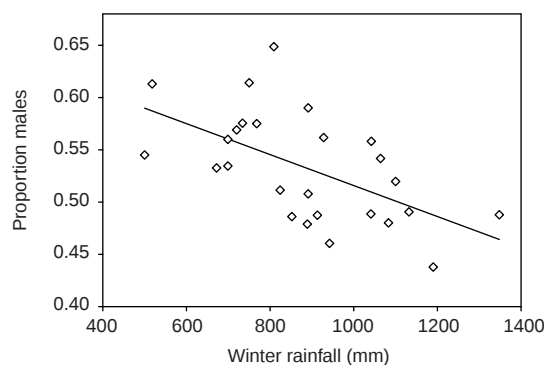


Figure 3 Proportion of males born each year in relation to winter rainfall (November to January, in mm), after correcting for the effect of density. Regression coefficient: $-0.0128 (\pm 0.046 \text{ s.e.})$, $t = -3.03$, d.f. = 23, $P = 0.006$; $r = -0.543$.

Table 2 Generalized linear mixed model of female fecundity

Female fecundity			
Term	Wald statistic (χ^2)	d.f.	P-value
Mother's age ²	5.7	2	0.058
Previous calf	221.2	1	<0.001
Mother's dominance	4.0	1	0.046
Winter rainfall	13.2	1	<0.001
Population density	12.5	1	<0.001
Dominance-density	0.2	1	0.66
Previous calf-density	9.3	1	0.002
Previous calf-rainfall	7.4	1	0.007
Density-rainfall	4.6	1	0.032
Minimal model	Average effect	s.e.	
Previous calf	-2.43	0.16	
Mother's dominance	0.38	0.19	
Winter rainfall	-0.0011	0.001	
Density	-0.00031	0.0058	
Previous calf-rainfall	-0.0015	0.00062	
Previous calf-density	-0.015	0.0059	
Density-rainfall	-0.000040	0.000019	

GLMM of female fecundity gives the probability that a female gives birth in a given year; $n = 2,269$. See legend to Table 1 for details.

plausible that the resorption of male fetuses by mothers that are unable to invest enough resources to provide them with a reasonable chance of breeding successfully is selectively advantageous⁴. However, an alternative explanation is that the high mortality of male fetuses is a by-product of faster male growth rates, which have evolved under sexual selection: in dimorphic species, males require more food and so will be more adversely affected by food restrictions²³. At least two lines of evidence support the latter suggestion for this population. Overall mortality rates are higher for juvenile males than females, and this differential in juvenile mortality increases with both population density²³ and winter rainfall²². It seems reasonable to suppose that male fetuses, like male calves, may be similarly more vulnerable than females.

As yet, no other study of wild mammals has allowed a comparison of both fecundity and offspring sex ratio between individuals under different environmental conditions. However, comparison with other ungulate studies indicates that there may be a link between resource availability and the tendency for females in good condition to produce male-biased sex ratios. In eight other ungulate studies, females in good condition produced a significantly higher proportion of male offspring than females in poorer condition in the same population. Of these populations, three were captive (pigs⁷, arrui⁶, sheep²⁴); two were semi-domesticated (reindeer⁸, bison⁹); one was in a game-management area (roe deer¹⁰); one was subject to exceptionally high predation rates (pronghorns, with almost 90% mortality in fawns²⁵); and the eighth involved red deer on Rum in the parts of the island where culling has been maintained¹¹. In all cases, population density was presumably below carrying capacity, or not limited by resource availability: this may be a necessary criterion for such associations to be detectable. It is, however, far from sufficient: studies that have found no association or an association in the opposite direction include a wide range of conditions from captivity^{12,14} and heavy culling¹³ to minimal culling²⁶, trophy hunting¹⁶ and a wild population²⁷.

The effect of population density on the sex ratio has implications for the management of red deer populations. Variation of up to 10% may seriously violate assumptions of equal birth rates employed in population dynamics models, and is likely to affect optimal harvesting rates and the growth rates of small populations. The result also highlights the need for adequate density control on sporting estates where red deer populations are managed to maximize the production of males.

The existence of more than one mechanism affecting birth sex ratios in a single population, and the dependence of these mechanisms on environmental conditions, could explain the inconsistency in sex-ratio deviations across mammalian populations. Adaptive

differences in the conception sex ratio related to individual phenotypic variation may require favourable environmental conditions, whereas differences in the susceptibility of male and female fetuses to nutritional stress may generate population-wide trends in annual birth sex ratios. In the Rum red deer population, the action of one mechanism swamped the other within about two generations. This may explain why general trends in sex-ratio variation have been so difficult to detect. □

Methods

Since 1971, life history data have been collected on individual red deer (*Cervus elaphus*) in the North Block of the Isle of Rum, Scotland, an area of about 12 km² (ref. 18). We take as our measure of density the number of females of more than one year old in the study area; since the cessation of culling in 1973, density has risen from 57 to 178. Females produce at most one calf per year throughout their breeding lifespan, whereas male reproductive success shows greater variance; adult male weight is 1.7 times that of females, and male calves are heavier at birth¹⁸. All animals in the study population are individually recognizable; daily monitoring of the population during the calving season shows whether or not each female calved in a given year, and, if so, the sex of her offspring. An age-corrected dominance rank for each female, ranging from 0 to 1, is calculated from observations of interactions between pairs of individuals, as described elsewhere.² Females are also classified as to whether or not they reared a calf the previous year that survived to six months. Average temperature and total rainfall in the following periods were considered: August to October, November to January, February to March. None of the weather variables showed consistent change over the study period.

We report two forms of statistical analysis. The proportion of males born each year (referred to as the annual birth sex ratio) was related to density and weather measures using simple linear regression (normality of errors was satisfied; 1973 was excluded owing to incomplete data collection). The probability that an individual calf was male and the probability that a female gave birth in a given year were analysed using generalized linear mixed models²⁸, with restricted maximum likelihood estimation of variance components. Just as generalized linear models allow the extension of general linear models to data where the errors are not normally distributed, GLMMs allow similar extensions to the conventional mixed model case where the response variable is determined by both random and fixed effects. In this case, the random component arose because of repeated sampling within a year and repeated sampling of the same females across years. Year and female identity were therefore fitted as random effects. In both models, the response variables were binary (male, not male; had calf, did not have calf), necessitating the use of a logit link function. The significance of the explanatory terms, the fixed effects, was assessed by their Wald statistics (distributed as χ^2) for each term when fitted last in the model. All interaction terms were tested, but are not reported unless statistically significant. Analysis was performed in Genstat 5, version 3.2.

Received 26 January; accepted 26 March 1999.

- Clutton-Brock, T. H. & Iason, G. R. Sex ratio variation in mammals. *Q. Rev. Biol.* **61**, 339–373 (1986).
- Clutton-Brock, T. H., Albon, S. D. & Guinness, F. E. Great expectations: dominance, breeding success and offspring sex ratios in red deer. *Anim. Behav.* **34**, 460–471 (1986).
- Clutton-Brock, T. H., Albon, S. D. & Guinness, F. E. Maternal dominance, breeding success and birth sex ratios in red deer. *Nature* **308**, 358–360 (1984).
- Trivers, R. L. & Willard, D. E. Natural selection of parental ability to vary the sex ratio of offspring. *Science* **179**, 90–92 (1973).
- Hiraiwa-Hasegawa, M. Skewed birth sex ratios in primates: should high-ranking mothers have daughters or sons? *Trends Ecol. Evol.* **8**, 395–399 (1993).
- Cassinello, J. High-ranking females bias their investment in favour of male calves in captive *Ammotragus iervia*. *Behav. Ecol. Sociobiol.* **38**, 417–424 (1996).
- Meikle, D. B., Drickamer, L. C., Bessey, S. H., Arthur, R. D. & Rosenthal, T. L. Dominance rank and parental investment in swine (*Sus scrofa domestica*). *Ethology* **102**, 969–978 (1996).
- Kojola, I. & Eloranta, E. Influences of maternal body weight, age, and parity on sex ratio in semidomesticated reindeer (*Rangifer t. tarandus*). *Evolution* **43**, 1331–1336 (1989).
- Rutberg, A. T. Lactation and fetal sex ratios in American bison. *Am. Nat.* **127**, 89–94 (1986).
- Wauters, L. A., Decrombrughe, S. A., Nour, N. & Matthysen, E. Do female roe deer in good condition produce more sons than daughters? *Behav. Ecol. Sociobiol.* **37**, 189–193 (1995).
- Flint, A. P. F., Albon, S. D. & Jafar, S. I. Blastocyst development and conceptus sex selection in red deer *Cervus elaphus*: studies of a free-living population on the Isle of Rum. *Gen. Comp. Endocrinol.* **106**, 374–383 (1997).
- Mendl, M., Zenlla, A. J., Broom, D. M. & Whittemore, C. T. Maternal social-status and birth sex-ratio in domestic pigs: an analysis of mechanisms. *Ethol. Sociobiol.* **16**, 257–333 (1995).
- Nygren, T. & Kojola, I. Twinning and fetal sex ratio in moose: effects of maternal age and mass. *Can. J. Zool.* **75**, 1945–1948 (1997).
- Birgesson, B. Adaptive adjustment of the sex ratio: more data and considerations from a fallow deer population. *Behav. Ecol.* **9**, 404–408 (1998).

- Hewison, A. J. M. & Gaillard, J. M. Birth-sex ratios and local resource competition in roe deer, *Capreolus capreolus*. *Behav. Ecol.* **7**, 461–464 (1996).
- Festa-Bianchet, M. The social system of bighorn sheep: grouping patterns, kinship and female dominance rank. *Anim. Behav.* **42**, 71–82 (1991).
- Albon, S. D., Coulson, T. N. & Clutton-Brock, T. H. In *Recent Developments in Deer Biology* (ed. Milne, J. A.) 85–95 (Macaulay Land Use Research Institute, Craigiebuckler, Aberdeen and Moredun Research Institute, Edinburgh, 1998).
- Clutton-Brock, T. H., Guinness, F. E. & Albon, S. D. *Red Deer—Behaviour and Ecology of Two Sexes* (University of Chicago Press, Chicago, 1982).
- Gomendio, M., Clutton-Brock, T. H., Albon, S. D., Guinness, F. E. & Simpson, M. J. Mammalian sex ratios and variation in costs of rearing sons and daughters. *Nature* **343**, 261–263 (1990).
- Flint, A. P. F., Albon, S. D., Loudon, A. S. I. & Jabbour, H. N. Behavioral dominance and corpus luteum function in red deer *Cervus elaphus*. *Horm. Behav.* **31**, 296–304 (1997).
- Albon, S. D., Clutton-Brock, T. H. & Guinness, F. E. Early development and population dynamics in red deer II. Density-independent effects and cohort variation. *J. Anim. Ecol.* **56**, 69–81 (1987).
- Coulson, T. N., Albon, S. D., Slate, J. & Pemberton, J. M. Sex dependent responses to inbreeding and outbreeding in red deer calves. *Evolution* (in the press).
- Clutton-Brock, T. H., Albon, S. D. & Guinness, F. E. Parental investment and sex differences in juvenile mortality in birds and mammals. *Nature* **313**, 131–133 (1985).
- Kent, J. P. Birth sex ratios in sheep over nine lambing seasons: years 7–9 and the effects of ageing. *Behav. Ecol. Sociobiol.* **36**, 101–104 (1995).
- Byers, J. A. *American Pronghorn: Social Adaptations and the Ghosts of Predators Past* (Univ. of Chicago Press, Chicago, 1997).
- Green, W. C. H. & Rothstein, A. Sex bias or equal opportunity—patterns of maternal investment in bison. *Behav. Ecol. Sociobiol.* **29**, 373–384 (1991).
- Lloyd, P. H. & Rasa, O. A. E. Status, reproductive success and fitness in Cape mountain zebra (*Equus zebra zebra*). *Behav. Ecol. Sociobiol.* **25**, 411–420 (1989).
- Schall, R. Estimation in generalized linear models with random effects. *Biometrika* **78**, 719–727 (1991).

Acknowledgements. We thank Scottish Natural Heritage for permission to work on Rum; their local staff for help and support; A. Alexander, A. Curnow, S. Morris and many others for field data collection; N. Barton, P. Brotherton, A. Cockburn, D. Elston, M. Forchhammer, H. Kruuk, J. Lindström, A. Manning and P. Meir for comments and discussion; and the NERC for funding this work.

Correspondence and requests for materials should be addressed to L.K. (e-mail: Loeske.Kruuk@ed.ac.uk).

Multifractality in human heartbeat dynamics

Plamen Ch. Ivanov*†, Luís A. Nunes Amaral*†, Ary L. Goldberger‡, Shlomo Havlin‡, Michael G. Rosenblum§, Zbigniew R. Struzik|| & H. Eugene Stanley*

* Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA

† Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA

‡ Gonda Goldschmid Center and Department of Physics, Bar-Ilan University, Ramat Gan, Israel

§ Department of Physics, Potsdam University, 14415 Potsdam, Germany

|| Centre for Mathematical and Computing Science, Kruislaan 413, NL-1098 SJ Amsterdam, The Netherlands

There is evidence that physiological signals under healthy conditions may have a fractal temporal structure¹. Here we investigate the possibility that time series generated by certain physiological control systems may be members of a special class of complex processes, termed multifractal, which require a large number of exponents to characterize their scaling properties^{2–6}. We report on evidence for multifractality in a biological dynamical system, the healthy human heartbeat, and show that the multifractal character and nonlinear properties of the healthy heart rate are encoded in the Fourier phases. We uncover a loss of multifractality for a life-threatening condition, congestive heart failure.

Biomedical signals are generated by complex self-regulating systems that process inputs with a broad range of characteristics^{7,8}. Many physiological time series, such as the one shown in Fig. 1a, are extremely inhomogeneous and non-stationary, fluctuating in an irregular and complex manner. The analysis of the fractal properties of such fluctuations has been restricted to second-order linear characteristics such as the power spectrum and the two-point autocorrelation function. These analyses reveal that the fractal behaviour of healthy, free-running physiological systems is often

characterized by $1/f$ -like scaling of the power spectra, $S(f)$, where f is the frequency^{9–12}.

Monofractal signals are homogeneous, in the sense that they have the same scaling properties throughout the entire signal. Therefore, monofractal signals can be indexed by a single global exponent—the Hurst exponent H (ref. 13). Multifractal signals, on the other hand, can be decomposed into many subsets characterized by different local Hurst exponents h , which quantify the local singular behaviour and thus relate to the local scaling of the time series. Thus, multifractal signals require many exponents to characterize their scaling properties fully^{4–6}.

The statistical properties of the different subsets characterized by these different exponents h can be quantified by the function $D(h)$, where $D(h_0)$ is the fractal dimension of the subset of the time series characterized by the local Hurst exponent h_0 ^{2,4–6}. Thus, the multifractal approach for signals, a concept introduced in the context of multi-affine functions^{14,15}, has the potential to describe a wide class of signals that are more complex than those characterized by a single fractal dimension (such as classical $1/f$ noise).

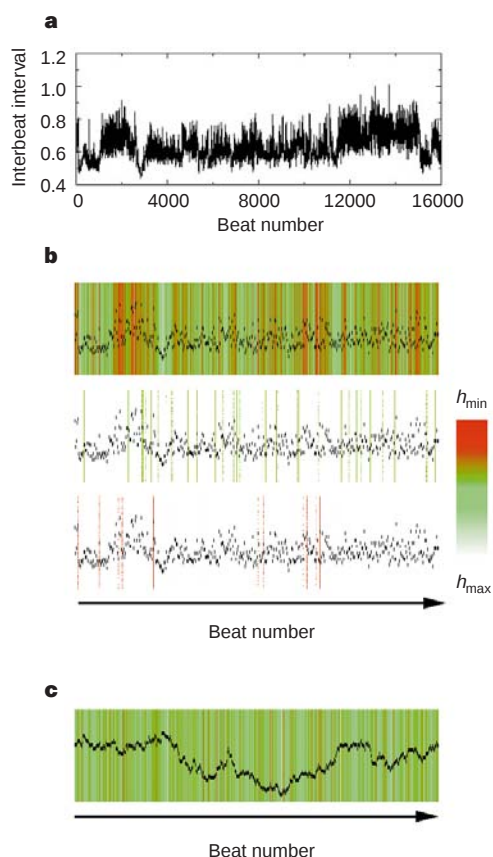


Figure 1 Visualization of multifractality in the heartbeat. **a**, Consecutive heartbeat intervals (in s) versus beat number for ~3-h data from a representative healthy subject. The time series exhibits very irregular and nonstationary behaviour. **b**, Top, the local Hurst exponents calculated for the 3-h record shown in **a**. The heterogeneity of the healthy heartbeat is represented by the broad range of local Hurst exponents h (colours) present and the complex temporal organization of the different exponents. Middle and bottom, the different fractal structures of two subsets of the time series characterized by different local Hurst exponents. The value of the local Hurst exponent for each subset is represented with a shade of green and red, respectively. The two subsets display different temporal structures which can be quantified by different fractal dimensions $D(h)$. **c**, The local Hurst exponents calculated for a monofractal signal—fractional Brownian motion with $H = 0.6$. The homogeneity of the signal is represented by the nearly monochromatic appearance of the signal which indicates that the local Hurst exponent h is the same throughout the signal and identical to the global Hurst exponent H .

We tested whether a large number of exponents is required to characterize heterogeneous heartbeat interval time series (Fig. 1) by undertaking multifractal analysis. The first problem is to extract the local value of h . To this end, we used methods derived from wavelet theory¹⁶. The properties of the wavelet transform make wavelet methods attractive for the analysis of complex non-stationary time series such as those found in physiological systems¹⁷. In particular, wavelets can remove polynomial trends that could cause box-counting techniques to fail to quantify the local scaling of the signal¹⁸. Additionally, the time-frequency localization properties of wavelets makes them particularly useful for revealing the underlying hierarchy that governs the temporal distribution of the local Hurst exponents¹⁹. Thus, the wavelet transform allows a reliable multifractal analysis to be performed^{18,19}. We used derivatives of the gaussian function as the analysing wavelet, which allowed us to estimate the singular behaviour and the corresponding exponent h at a given location in the time series. The higher the order, n , of the derivative, the higher the order of the polynomial trends removed and the better the detection of the temporal structure of the local scaling exponents in the signal.

We evaluated the local exponent h through the modulus of the maxima values of the wavelet transform at each point in the time series. We then estimated the scaling of the partition function $Z_q(a)$, which is defined as the sum of the q th powers of the local maxima of the modulus of the wavelet transform coefficients at scale a (ref. 19). For small scales, we expect

$$Z_q(a) \approx a^{\tau(q)}. \quad (1)$$

For certain values of q , the exponents $\tau(q)$ have familiar meanings. In particular, $\tau(2)$ is related to the scaling exponent of the Fourier power spectra, $S(f) \sim 1/f^\beta$, as $\beta = 2 + \tau(2)$. For positive q , $Z_q(a)$ reflects the scaling of the large fluctuations and strong singularities, whereas for negative q , $Z_q(a)$ reflects the scaling of the small fluctuations and weak singularities^{4,5}. Thus, the scaling exponents $\tau(q)$ can reveal different aspects of cardiac dynamics.

Monofractal signals display a linear $\tau(q)$ spectrum, $\tau(q) = qH - 1$, where H is the global Hurst exponent. For multifractal signals, $\tau(q)$ is a nonlinear function: $\tau(q) = qh - D(h)$, where $h = d\tau/dq$ is not constant. The fractal dimension $D(h)$, introduced earlier, is related to $\tau(q)$ through a Legendre transform:

$$D(h) = qh - \tau(q). \quad (2)$$

We analysed both daytime (12:00 to 18:00) and night-time (0:00 to 6:00) heartbeat time series records from healthy subjects, and the daytime records of patients with congestive heart failure. These data were obtained by Holter monitoring²⁰. Our database includes 18 healthy subjects (13 female and 5 male, with ages between 20 and 50, average 34.3 years), and 12 congestive heart failure subjects (3 female and 9 male, with ages between 22 and 71, average 60.8 years) in sinus rhythm (see Methods for details on data acquisition and preprocessing). For all subjects, for a broad range of positive and negative q , the partition function $Z_q(a)$ scales as a power law (Fig. 2a, b).

For all healthy subjects, $\tau(q)$ is a nonlinear function (Figs 2c and 3a), which indicates that the heart rate of healthy humans is a multifractal signal. Figure 3b shows that, for healthy subjects, $D(h)$ has non-zero values for a broad range of local Hurst exponents h . The multifractality of healthy heartbeat dynamics cannot be explained by activity, as we analyse data from subjects at night. Furthermore, this multifractal behaviour cannot be attributed to sleep-stage transitions, as we find multifractal features during daytime hours as well. The range of scaling exponents ($0 < h < 0.3$) with non-zero fractal dimension $D(h)$ indicates that the fluctuations in the healthy heartbeat dynamics exhibit anti-correlated behaviour ($h = 1/2$ corresponds to uncorrelated behaviour; $h > 1/2$ corresponds to correlated behaviour).

In contrast, heart-rate data from subjects with a pathological

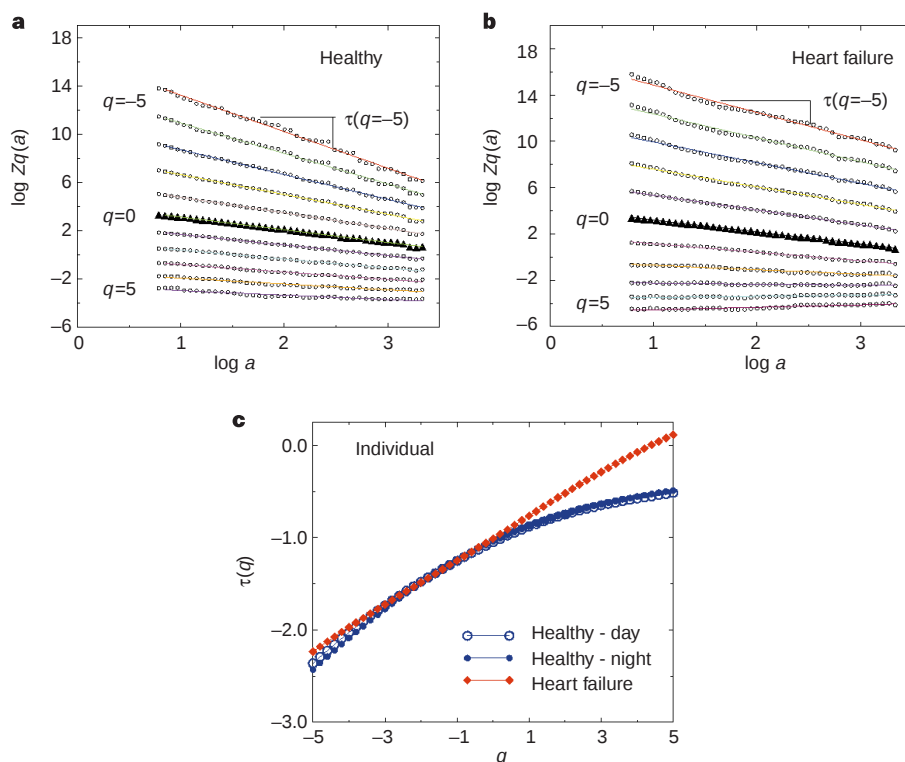


Figure 2 Heartbeat time series contain densely packed, non-isolated singularities which unavoidably affect each other in the time-frequency decomposition. Therefore, rather than evaluating the distribution of the inherently unstable local singularity exponents (as shown in Fig. 1), we estimate the scaling of an appropriately chosen global measure: the q moments of the probability distribution of the maxima of the wavelet transform $Z_q(a)$ (using the third derivative of the gaussian function as the analysing wavelet). The scaling of the partition function $Z_q(a)$ with scale a was obtained from daytime records consisting of $\sim 25,000$ beats for **a**, a healthy subject, and **b**, a subject with congestive heart failure. We calculate $\tau(q)$ for moments $q = -5, -4, \dots, 0, \dots, 5$ and scales $a = 2 \times 1.15^i, i = 0, \dots, 41$. We

show the calculated values of $Z_q(a)$ for scales $a > 8$. Top curve $q = -5$; middle curve (bold), $q = 0$; bottom curve, $q = 5$. $\tau(q)$ are obtained from the slope of the curves in the region $16 < a < 700$, thus eliminating the influence of any residual small-scale random noise due to signal pre-processing as well as extreme, large-scale fluctuations of the signal. **c**, Multifractal spectrum $\tau(q)$ for individual records. A monofractal signal would correspond to a straight line for $\tau(q)$; for a multifractal signal $\tau(q)$ is nonlinear. Note the differences between the curves for healthy and heart-failure records. The changing curvature for the healthy records indicates multifractality. In contrast, $\tau(q)$ is linear for the subject with congestive heart failure, indicating monofractality.

condition—congestive heart failure—show a clear loss of multifractality (Fig. 3a, b). For the heart-failure subjects, $\tau(q)$ is close to linear and $D(h)$ is nonzero only over a very narrow range of exponents h , indicating monofractal behaviour (Fig. 3).

Our results show that, for healthy subjects, local Hurst exponents in the range $0.07 < h < 0.17$ are associated with fractal dimensions close to one. This means that the subsets characterized by these local exponents are statistically dominant. On the other hand, for the

heart-failure subjects, the statistically dominant exponents are confined to a narrow range of local Hurst exponents centred at $h \approx 0.22$. These results suggest that, for heart failure, the fluctuations are less anti-correlated than for healthy dynamics, as the dominant scaling exponents h are closer to $1/2$.

We systematically compared our method with other widely used methods of heart-rate time-series analysis^{12,21,22}. Several of these methods do not result in a fully consistent assignment of healthy

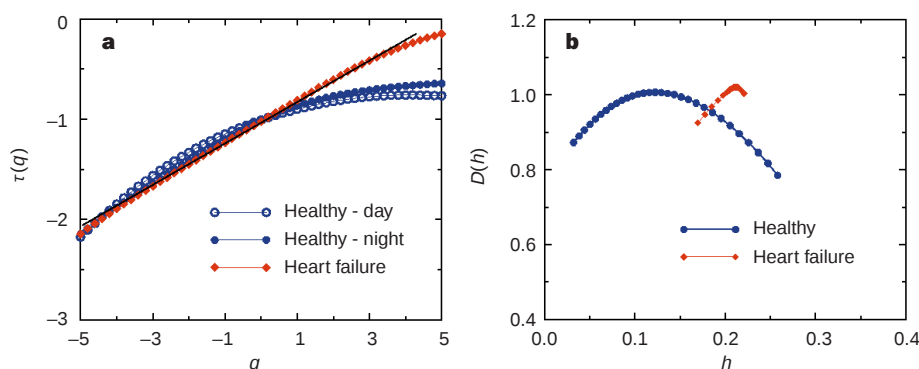


Figure 3 Multifractality in healthy dynamics. **a**, Multifractal spectrum $\tau(q)$ of the group averages for daytime and night-time records for 18 healthy subjects and 12 patients with congestive heart failure. The results show multifractal behaviour for the healthy group and different behaviour for the heart-failure group. **b**, Fractal dimensions $D(h)$ obtained through a Legendre transform from the group-

averaged $\tau(q)$ spectra of **a**. The shape of $D(h)$ for the individual records and for the group average is broad, indicating multifractal behaviour. $D(h)$ for the heart-failure group is very narrow, indicating monofractality. The different form of $D(h)$ for the heart-failure group may reflect perturbation of the cardiac neuroautonomic control mechanisms associated with this pathology.

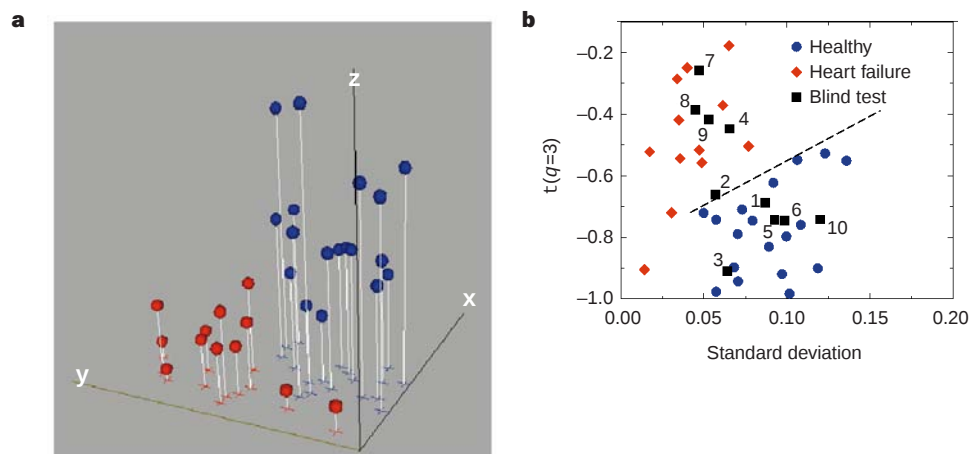


Figure 4 Discrimination method based on the multifractal formalism. **a**, Each subject's dataset in the database is characterized by three quantities. The first quantity (z -axis) is the degree of multifractality, which is the difference between the maximum and minimum values of local Hurst exponent h for each individual. The degree of multifractality is zero for a monofractal. The second quantity (y -axis) is the exponent value $\tau(q=3)$ characterizing the scaling of the third moment, $Z_3(a)$. The third quantity (x -axis) is the standard deviation of the interbeat intervals. Blue spheres, healthy subjects; red spheres, heart-failure subjects. This multifractal approach discriminates healthy from heart-failure subjects. **b**, Discrimination method based on multifractal formalism for the 'blind' datasets. The y -axis is the exponent value $\tau(q=3)$, and the x -axis is the standard deviation of the time series of interbeat intervals. Black squares show the results for the datasets of the blind test. Two of the heart-failure subjects (closest to the origin) do not fit as clearly in either group; however, they are clearly identified as unhealthy in **a** because their degree of multifractality (z -axis) is close to zero (these two subjects are indicated by the red spheres in **a** that are closest to the origin). These cases demonstrate that a single exponent is not sufficient to describe the complexity of a heartbeat time series.

versus diseased subjects (see <http://polymer.bu.edu/~amaral/Heart.html>). Figure 4a shows the results of our method based on the multifractal formalism. Each subject's dataset is characterized by three quantities: (1) the standard deviation of the interbeat intervals; (2) the exponent value $\tau(q=3)$ obtained from the scaling of the third moment $Z_3(a)$; and (3) the degree of multifractality, defined as the difference between the maximum and minimum values of local Hurst exponent h for each individual (Fig. 5). The multifractal approach robustly discriminates healthy subjects from heart-failure subjects.

We next blindly analysed a separate database containing 10 records, 5 from healthy individuals and 5 from patients with congestive heart failure. The time series in the new database are shorter than those in our database; on average they are only 2-h long (less than 8,000 beats). Figure 4b shows the projection on the x - y plane of our data presented in Fig. 4a. The results for the blind test

are shown in black. Our approach clearly separates the blind test subjects into two groups: 1, 3, 5, 6 and 10 fall in the healthy group, and 2, 4, 7, 8 and 9 in the heart-failure group. Unblinding the test code confirms these assignments. Thus, an analysis incorporating the multifractal method may add diagnostic power to contemporary analytic methods of heartbeat (and other physiological) time-series analysis.

The multifractality of heart-beat time series also enables us to quantify the greater complexity of the healthy dynamics compared to those of pathological conditions. Power-spectrum analysis defines the complexity of heart-beat dynamics through its scale-free behaviour, identifying a single scaling exponent as an index of healthy or pathological behaviour. Hence, the power spectrum cannot quantify the greater level of complexity of the healthy dynamics, reflected in the heterogeneity of the signal. In contrast, the multifractal analysis reveals this new level of complexity by the

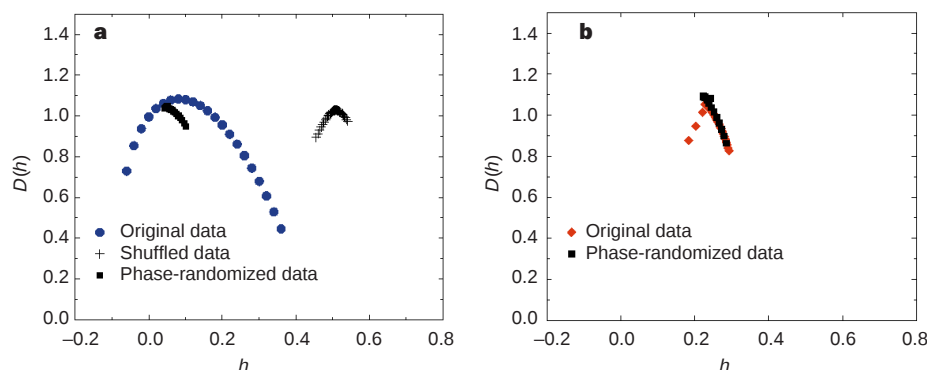


Figure 5 Nonlinearity and Fourier phases. **a**, The fractal dimensions $D(h)$ for a 6-h daytime record from a healthy subject. After reshuffling and integrating the increments in this interbeat-interval time series, so that all correlations are lost but the distribution is preserved, we obtain monofractal behaviour—a very narrow point-like spectrum centred at $h \equiv H = 1/2$. Such behaviour corresponds to a simple random walk. A different test, in which the $1/f$ scaling of the heartbeat signal is preserved but the Fourier phases are randomized (that is, nonlinearities

are eliminated), leads again to a monofractal spectrum centred at $h = 0.07$, because the linear correlations were preserved. These tests indicate that the multifractality is related to nonlinear features of the healthy heartbeat dynamics, rather than to the ordering or the distribution of the interbeat intervals in the time series. **b**, The fractal dimensions $D(h)$ for a 6-h daytime record from a heart-failure subject. The narrow multifractal spectrum indicates loss of multifractal complexity and reduction of nonlinearities with pathology.

broad range of exponents necessary to characterize the healthy dynamics. Moreover, the change in shape of the $D(h)$ curve for the heart-failure group may provide insight into the changes in the cardiac control mechanisms resulting from this pathology.

To study the complexity of the healthy dynamics further, we performed two tests with surrogate time series. First, we generated a surrogate time series by shuffling the interbeat interval increments of a record from a healthy subject. The new signal preserves the distribution of interbeat interval increments but destroys the long-range correlations among them. Hence, the signal is a simple random walk, which is characterized by a single Hurst exponent $H = 1/2$ and exhibits monofractal behaviour (Fig. 5a). Second, we generated a surrogate time series by performing a Fourier transform on a record from a healthy subject, preserving the amplitudes of the Fourier transform but randomizing the phases, and then performing an inverse Fourier transform. This procedure eliminates nonlinearities, preserving only the linear features of the original time series. The new surrogate signal has the same $1/f$ behaviour in the power spectrum as the original heart-beat time series; however, it exhibits monofractal behaviour (Fig. 5a). We repeated this test on a record from a heart-failure subject. In this case, there is a smaller change in the multifractal spectrum (Fig. 5b). The results suggest that the healthy heartbeat time series contains important phase correlations which are cancelled in the surrogate signal by the randomization of the Fourier phases, and that these correlations are weaker in heart-failure subjects. Furthermore, our analysis indicates that the observed multifractality is related to nonlinear features of the healthy heartbeat dynamics. Several studies have tested for nonlinear and deterministic properties in records of interbeat intervals^{23–27}. We have demonstrated an explicit relation between the nonlinear features (represented by the Fourier phase interactions) and the multifractality of healthy cardiac dynamics (Fig. 5).

From a physiological perspective, the detection of robust multifractal scaling in the heart-rate dynamics is of interest because it indicates that the control mechanisms regulating the heartbeat might interact as part of a coupled cascade of feedback loops in a system operating far from equilibrium^{28,29}. Furthermore, these results indicate that the healthy heartbeat is even more complex than previously suspected, posing a challenge to ongoing efforts to develop realistic models of the control of heart rate and other processes under neuroautonomic regulation. □

Methods

Heart failure (CHF) data were recorded using a Del Mar Avionics Model 445 Holter recorder and digitized at 250 Hz. Beats were labelled using 'Aristotle' arrhythmia analysis software, which labels each detected beat as normal, ventricular ectopic, supraventricular ectopic or unknown³⁰. The location of the R-wave peaks is determined with a resolution of 4 ms.

Healthy datasets were recorded using a Marquette Electronics series 8500 Holter recorder. Using a Marquette Electronics model 8000T Holter scanner, the tapes were then digitized at 128 Hz, scanned and annotated. The annotations were manually verified by an experienced Holter scanning technician. The location of the R-wave peaks was thus determined with a resolution of 8 ms.

The finite resolution implies that our estimates of the interbeat intervals are affected by a white noise due to estimation error. The signal-to-noise ratio for both healthy and heart failure is of the order of 100. Furthermore, the white noise due to the measurements would lead to the detection of a local Hurst exponent $h = 0$ at very small scales. For that reason, we have considered only scales larger than 16 beats.

From the beat annotation file, only the intervals (NN) between consecutive normal beats were determined; thus intervals containing non-normal beats were eliminated from the NN interval series. For the CHF data, an average of 2% (range, 0.1 to 0.6%) of the intervals were eliminated, and for the normal data an average of 0.01% (range 0 to 0.06%) were eliminated. No interpolation was done for eliminated intervals.

To eliminate outliers due to missed beat detections which would give rise to erroneously large intervals that may have been included in the NN interval series, a moving-window average filter was applied. For each set of five contiguous NN intervals, a local mean was computed, excluding the central interval. If the value of the central interval was greater than twice the local average, it was considered to be an outlier and excluded from the NN interval series. This criterion was applied to each NN interval in the series. For the CHF data, an average of 0.02% (range 0 to 0.1%) of the intervals were eliminated, and for the normal data an average of 0.07% (range 0 to 0.7%) were eliminated. No interpolation was done for eliminated intervals. Overall a total of 2% (range, 0.1 to 6%) of the total number of NN intervals were eliminated for the CHF data and 0.08% (range 0 min to 0.7% max) for the normal data.

Next, we built a time series of increments between consecutive NN intervals and calculated their standard deviation. We then identified all pairs of associated increments with opposite signs and with an amplitude larger than 3 s.d. The values of the increments for each pair were replaced by linear interpolations of their values and the time series of NN interval time was reconstructed by integration from the filtered increments. About 1% of time intervals were corrected by this procedure.

Received 2 March; accepted 7 April 1999.

1. Bassingthwaite, J. B., Liebovitch, L. S. & West, B. J. *Fractal Physiology* (Oxford Univ. Press, New York, 1994).
2. Dewey, T. G. *Fractals in Molecular Biophysics* (Oxford Univ. Press, Oxford, 1997).
3. Stanley, H. E. & Meakin, P. Multifractal phenomena in physics and chemistry. *Nature* **335**, 405–409 (1988).
4. Vicsek, T. *Fractal Growth Phenomena*, 2nd edn (World Scientific, Singapore, 1993).
5. Takayasu, H. *Fractals in the Physical Sciences* (Manchester Univ. Press, Manchester, UK, 1997).
6. Stanley, H. E. in *Fractals and Disordered Systems* 2nd edn (eds Bunde, A. & Havlin, S.) 1–68 (Springer, Berlin, 1996).
7. Shlesinger, M. F. Fractal time and $1/f$ noise in complex systems. *Ann. NY Acad. Sci.* **504**, 214–228 (1987).
8. Malik, M. & Camm, A. J. (eds) *Heart Rate Variability* (Futura, Armonk, NY, 1995).
9. Akselrod, S. et al. Power spectrum analysis of heart rate fluctuations: a quantitative probe of beat-to-beat cardiovascular control. *Science* **213**, 220–222 (1981).
10. Kobayashi, M. & Musha, T. $1/f$ fluctuation of heartbeat period. *IEEE Trans. Biomed. Eng.* **29**, 456–457 (1982).
11. Hausdorff, J. M. et al. Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations. *J. Appl. Physiol.* **80**, 1448–1457 (1996).
12. Peng, C.-K., Havlin, S., Stanley, H. E. & Goldberger, A. L. Quantification of scaling exponents and crossover phenomena in nonstationary time series. *Chaos* **5**, 82–87 (1995).
13. Hurst, H. E. Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civ. Eng.* **116**, 770–808 (1951).
14. Vicsek, T. & Barabási, A. L. Multi-affine model for the velocity distribution in fully turbulent flows. *J. Phys. A* **24**, L845–L851 (1991).
15. Barabási, A.-L. & Stanley, H. E. *Fractal Concepts in Surface Growth* Ch. 24 (Cambridge Univ. Press, Cambridge, 1995).
16. Daubechies, I. *Ten Lectures on Wavelets* (SIAM, Philadelphia, 1992).
17. Ivanov, P. Ch. et al. Scaling behaviour of heartbeat intervals obtained by wavelet-based time-series analysis. *Nature* **383**, 323–327 (1996).
18. Muzy, J. F., Bacry, E. & Arneodo, A. Wavelets and multifractal formalism for singular signals: application to turbulence data. *Phys. Rev. Lett.* **67**, 3515–3518 (1991).
19. Muzy, J. F., Bacry, E. & Arneodo, A. The multifractal formalism revisited with wavelets. *Int. J. Bifurc. Chaos* **4**, 245–302 (1994).
20. Peng, C.-K. et al. Fractal mechanisms and heart rate dynamics: Long-range correlations and their breakdown with disease. *J. Electrocardiol.* **28**, 59–65 (1996).
21. Thurner, S., Feurstein, M. C. & Teich, M. C. Multiresolution wavelet analysis of heartbeat intervals discriminates healthy patients from those with cardiac pathology. *Phys. Rev. Lett.* **80**, 1544–2391 (1998).
22. Amaral, L. A. N., Goldberger, A. L., Ivanov, P. Ch. & Stanley, H. E. Scale-independent measures and pathologic cardiac dynamics. *Phys. Rev. Lett.* **81**, 2388–2391 (1998).
23. Lefebvre, J. et al. Predictability of normal heart rhythms and deterministic chaos. *Chaos* **3**, 267–276 (1993).
24. Yamamoto, Y. et al. Operation Everest II: an indication of deterministic chaos in human heart rate variability at simulated extreme altitude. *Biol. Cybern.* **69**, 205–212 (1993).
25. Kanters, J. K., Holstein-Rathlou, N. H. & Agner, E. Lack of evidence for low-dimensional chaos in heart rate variability. *J. Cardiovasc. Electrophysiol.* **5**, 128–137 (1994).
26. Sugihara, G., Allan, W., Sobel, D. & Allan, K. D. Nonlinear control of heart rate variability in human infants. *Proc. Natl Acad. Sci. USA* **93**, 2608–2613 (1996).
27. Poon, C.-S. & Merrill, K. K. Decrease of cardiac chaos in congestive heart failure. *Nature* **389**, 492–495 (1997).
28. Meneveau, C. & Sreenivasan, K. R. Simple multifractal cascade model for fully developed turbulence. *Phys. Rev. Lett.* **59**, 1424–1427 (1987).
29. Ivanov, P. Ch., Amaral, L. A. N., Goldberger, A. L. & Stanley, H. E. Stochastic feedback and the regulation of biological rhythms. *Europhys. Lett.* **43**, 363–368 (1998).
30. Moody, G. B. & Mark, R. G. Development and evaluation of a 2-lead ECG analysis program. *Computers Cardiol.* **9**, 39–44 (1983).

Acknowledgements. We thank A. Bunde, U. Frisch, J. M. Hausdorff, V. Horváth, H. Kallabis, R. G. Mark, J. Mietus, C.-K. Peng, K. R. Sreenivasan and B. J. West for discussions. We are especially grateful to A. Arneodo and T. Vicsek for advice on the analytical technique and on the text and R. Goldsmith for providing the blind test data. This work was supported by NASA, NIH, FCT/Portugal, and The G. Harold and Leila Y. Mathers Charitable Foundation.

Correspondence and requests for materials should be addressed to P.Ch.I. (e-mail: plamen@buphy.bu.edu)

Decrystallization of adult birdsong by perturbation of auditory feedback

Anthony Leonardo & Masakazu Konishi

Computation and Neural Systems Program and Division of Biology, MC 216-76, California Institute of Technology, Pasadena, California 91125, USA

Young birds learn to sing by using auditory feedback to compare their own vocalizations to a memorized or innate song pattern; if they are deafened as juveniles, they will not develop normal songs^{1,2}. The completion of song development is called crystallization. After this stage, song shows little variation in its temporal or spectral properties. However, the mechanisms underlying this stability are largely unknown. Here we present evidence that auditory feedback is actively used in adulthood to maintain the stability of song structure. We found that perturbing auditory feedback during singing in adult zebra finches caused their song to deteriorate slowly. This 'decrystallization' consisted of a marked loss of the spectral and temporal stereotypy seen in crystallized song, including stuttering, creation, deletion and distortion of song syllables. After normal feedback was restored, these deviations gradually disappeared and the original song was recovered. Thus, adult birds that do not learn new songs never-

theless retain a significant amount of plasticity in the brain.

The song of zebra finches consists of three levels of organization: syllables, which are individual sound components of the song separated by silent intervals; motifs, which are sequences of syllables; and bouts, which are sequences of motifs³. The spectral structure of the syllables is unstable in the juvenile bird. Similarly, motifs and bouts are organized differently from song to song. However, as the song gradually assumes its adult form, these different levels of organization become highly stereotyped, the variability of the spectral structure of the syllables becomes extremely small, and the bird sings these syllables in a highly predictable order. At this stage, the song is referred to as crystallized. Some birds, like the zebra finch, maintain their crystallized song throughout adulthood and are called age-limited learners. Open-ended learners like canaries can, in contrast, learn new songs in adulthood⁴.

The stability of song in age-limited learners was previously thought to be maintained without auditory feedback^{1,2}. Recent reports, however, show that deafening these birds after crystallization causes changes in song, suggesting that some auditory feedback is important for song maintenance throughout life⁵⁻⁷. Six to eight weeks after deafening, the song of adult zebra finches deteriorates, and shows addition and deletion of syllables, abnormal repetition of syllables (stuttering), and modified syllable sequences. By opening the auditory feedback loop, deafening shows how well the song pattern generator can maintain its original output without this signal. However, to learn how the song control system works, it is necessary to manipulate auditory feedback without disabling

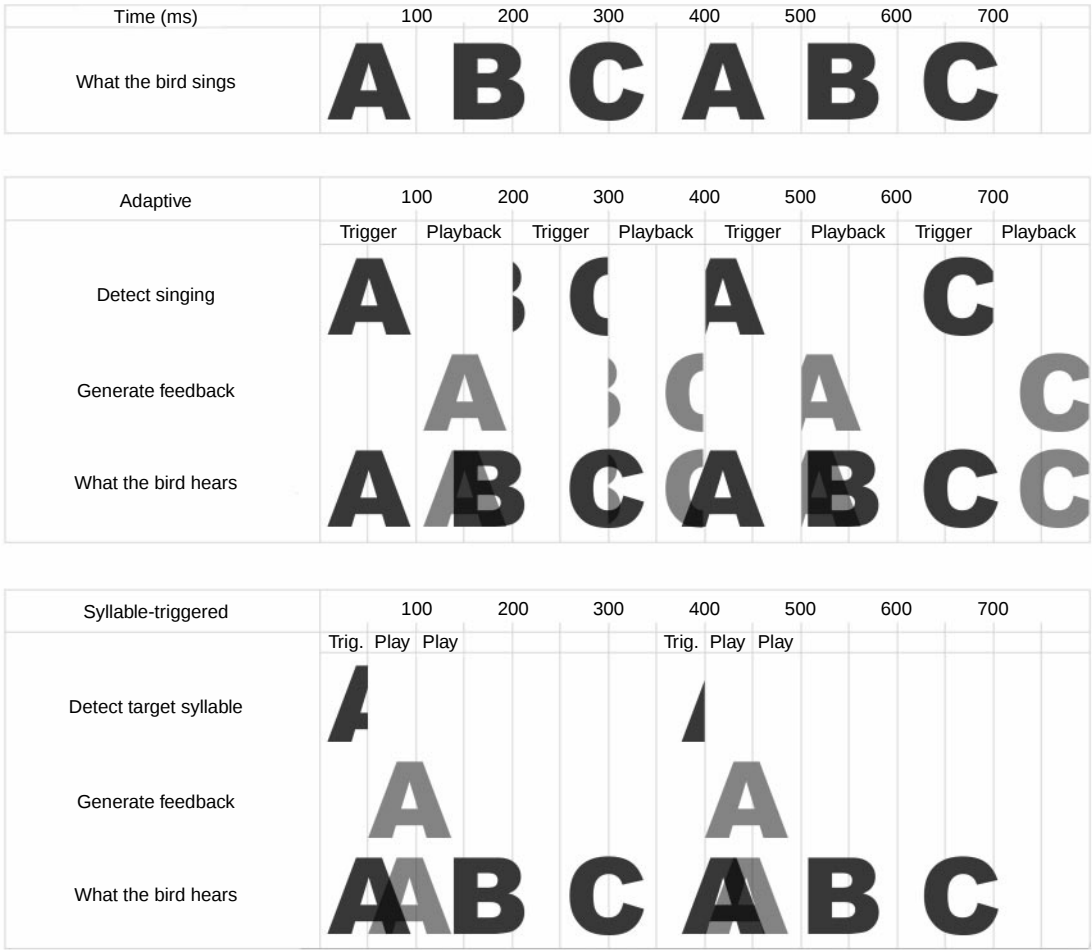


Figure 1 Protocols for constructing the feedback signals. In the adaptive protocol, the computer alternated between recording vocalizations and playing the last vocalization back to the bird. In the syllable-triggered protocol, the computer

monitored the bird's song in 50-ms bins and detected the production of a single target syllable. In both protocols, the bird heard the superposition of his own vocalizations and the computer-generated feedback.

We developed a computer-controlled system to perturb auditory feedback and designed two feedback paradigms (Fig. 1). In both methods, the computer detected singing and then played a feedback signal to the bird. Because the playback signals were delivered through an overhead speaker, the birds heard a superposition of natural and artificial sounds. In the first paradigm, the adaptive protocol ($n = 3$ birds), a computer alternated between recording

We recorded the songs of five adult male zebra finches in separate, sound-attenuated chambers for several weeks before beginning the experiment. The variability of their songs was well within that found in crystallized songs. The birds were then placed in the feedback system. After 1–4 months in this environment, four of the five birds



decrystallized song of the bird shown in **a**. The syllable sequences now vary from motif to motif, indicating a loss of the fundamental temporal stereotypy which characterizes crystallized zebra finch song. Motifs one and three contain new song syllables (F, G, H) which were not present in the baseline motif shown in **a**. Motif two contains a repetition of syllable 'B'. After the feedback was removed, the sequences shown in **b** and **c** gradually disappeared and only the baseline song was produced.

showed dramatic changes in their songs. To study the progression of these changes, we allowed some of the birds to sing without artificial feedback on a randomly chosen 10–15% of their song deliveries one day per week. All of the analysed data consisted of recordings of the bird singing by himself, with no artificial feedback. After the feedback was permanently stopped, we tracked all the birds for another 8–16 months to determine whether they could recover their original songs.

Decrystallization of the song occurred in both global song organization and local spectral structure, and consisted of the emergence of new spectral and temporal properties and increased occurrence of properties that were rare in the baseline song. We use

the term decrystallization to refer collectively to all of the perturbation-induced changes from the original quantitative and statistical structure of the song; this does not necessarily imply a return to a juvenile state of song structure. These changes include stuttering, creation, deletion and distortion of song syllables. As the decrystallization progressed, the proportion of normal songs decreased and that of abnormal ones increased. However, baseline songs were still produced with low probability even at the peak of song degradation.

The changes seen in the three adaptive protocol birds were very similar to those seen in deafened birds^{5–7}. A series of spectrograms representative of these results is shown in Fig. 2. Stuttering occurred in all three adaptive-protocol birds and was the most dramatic change in song organization. Both complex syllables and modified introductory notes were stuttered. A secondary effect of stuttering was a substantial increase in the maximum song length. Other changes to song organization induced by the feedback were the addition of new syllables to the song and, infrequently, the deletion of old syllables from the song. Many zebra finch syllables contain sets of frequencies that are integer multiples of a common fundamental frequency; such sets are called harmonic stacks (Fig. 2a, syllable D). The same three birds also showed spectral distortion in their song syllables, including wobbles in the harmonic structure of simple notes and the production of two superimposed harmonic stacks, indicating a loss of precise control over the vocal organ (the syrinx)^{8,9}. There was considerable variability in the magnitude and time course of the changes between different birds, but significant changes were generally seen within six weeks. Finally, all the changes in song structure described above could occur within different motifs in the same bout. This is significant because one of the hallmarks of crystallized song is its robust temporal stereotypy—an identical syllable sequence is maintained within all the motifs of a bout. Decrystallized song, in contrast, lacks this stereotypy (Fig. 2c).

Two birds received feedback in the syllable-triggered protocol. In one of the birds, significant changes in the spectrum of the target syllable appeared in less than a week and increased in magnitude for the remainder of the feedback period. No changes were seen in the spectrum or the ordering of any of the other syllables in the song. The changes to the target syllable consisted of the appearance of harmonic frequencies around previously single-frequency portions of the syllable's spectrum (Fig. 3). The presence of these additional harmonics grew more frequent with time, until eventually this initially tonal syllable was sometimes produced as a distorted harmonic stack. Humans can also alter the spectral structure of a sound in response to altered auditory feedback¹⁰. The second bird used in this protocol showed no changes in his song after 11 weeks of feedback perturbation.

Decrystallization essentially consisted of a large increase in the variability of the song. To quantify the changes associated with the arrangement and variability of syllable sequences, we developed an automatic method to sort the data acquired on a given day into syllable types by using the different spectral and temporal features of each syllable. This transformed the raw voltage waveforms recorded from the microphone into a series of syllable strings, such as 'A B C D E', from which we calculated the probability of different syllable sequences. We define the baseline motif as the most probable sequence of syllables that the bird repeated in a bout before the feedback period began. Figure 4a shows the probability of singing the baseline motif as a function of time for one bird. After receiving one month of perturbed feedback, the probability of the baseline motif for this bird decreased by a factor of six.

The probabilistic sequencing of syllables by the bird on a particular day can be fully characterized as a Markov chain¹¹. Thus, the likelihood of singing a particular syllable depends only on the occurrence of the last syllable produced, and not on any prior syllables. The bird's song over the course of the experiment can then be described as a Markov chain which evolves in time (Fig. 4a), and

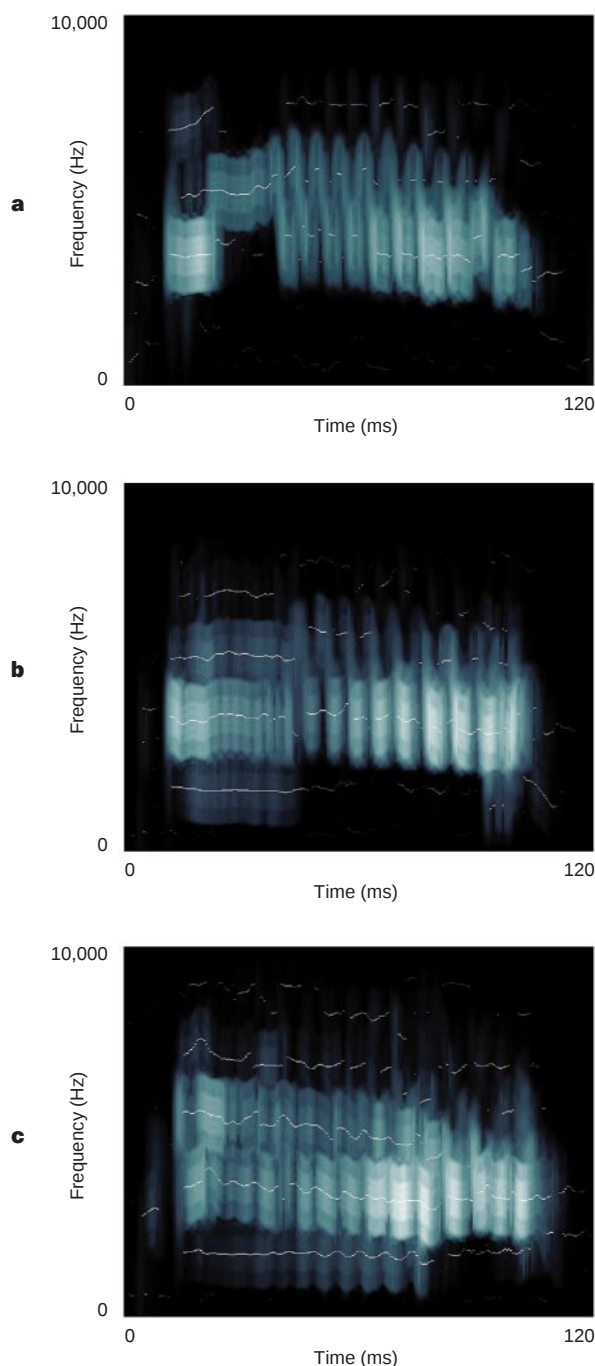


Figure 3 Decrystallization of a single syllable. **a**, Baseline (pre-feedback) version of the syllable. **b**, After one week of syllable-triggered feedback. Harmonic frequencies (white lines) have appeared around the single frequency in the early portion of the syllable (at $t \approx 30$ ms). **c**, After one month of feedback, additional harmonic frequencies now stretch throughout the duration of the syllable.

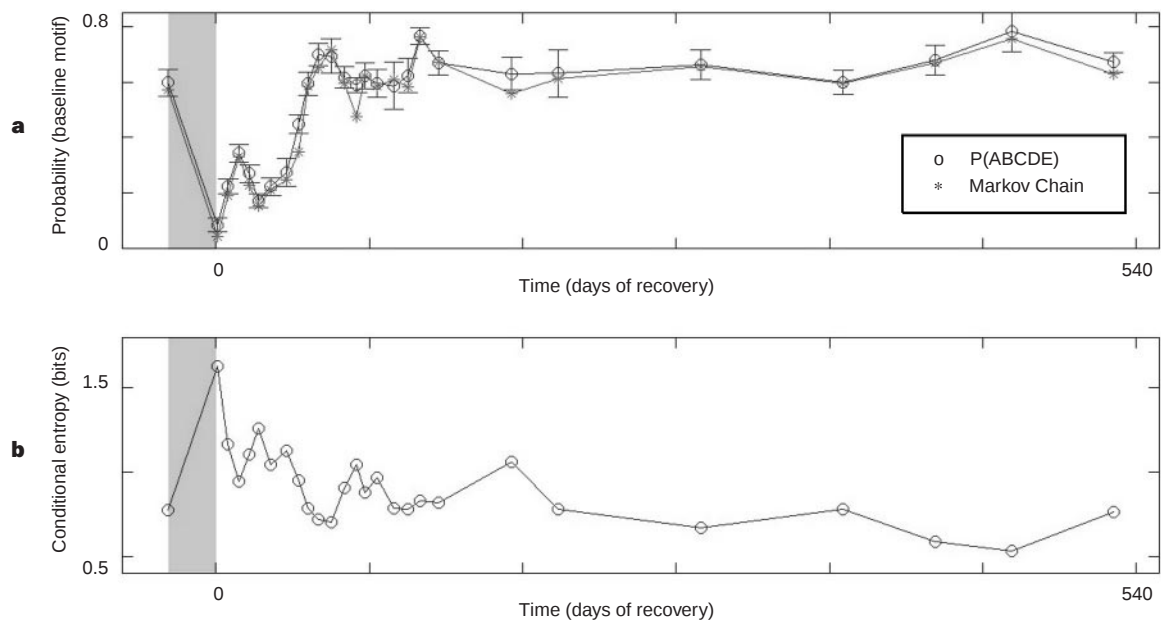


Figure 4 Time course of decrystallization and recovery. **a**, The probability of one bird singing his baseline motif as a function of time. The shaded area represents the feedback interval. Circles represent the probability calculated from the entire sequence of syllables (Prob {A B C D E}); asterisks represent the probability

calculated from the syllable-to-syllable transition probabilities of the Markov chain for the song ($\text{Prob}\{A\} \cdot \text{Prob}\{B|A\} \cdot \text{Prob}\{C|B\} \cdot \text{Prob}\{D|C\} \cdot \text{Prob}\{E|D\}$). The two curves are identical to within 4% error. **b**, The conditional entropy of the song, which is a measure of song variability, for the same bird.

the conditional entropy of the Markov chain can be used as an estimate of the variability of the song on a particular day (Fig. 4b)¹². The conditional entropy measures the uncertainty in observing syllable B, given that syllable A was just produced. If each syllable is followed only by a single other syllable (only one type of syllable sequence is produced), then the conditional entropy will be low, whereas if any syllable can follow any other syllable (many different sequences are produced), the conditional entropy will be high. We examined the data for all three adaptive protocol birds using this method and found that birds in a decrystallized state had a significantly higher conditional entropy than they did in their baseline state. Taken together, the time course of the baseline motif probability and the conditional entropy shows that the bird in Fig. 4 went from being a low-variance singer, with essentially a single motif, to being a high-variance singer who produced a large number of different motifs.

After the removal of the artificial feedback, the temporal and spectral variability in the songs of all three adaptive protocol birds slowly decreased over the following weeks and months. The songs eventually recrystallized and became stable again. Both the probability of the baseline motif and the conditional entropy of the song returned to their baseline levels. Stuttering, abnormal sequencing of syllables and modified spectral organization gradually became infrequent and were replaced by the temporal and spectral organizations characteristic of the original song. A complete recovery took about 2–4 months. The syllable-triggered bird made a partial recovery by 8 months after the cessation of feedback. The slow progression of these changes suggests that auditory feedback does not exert a great deal of instantaneous control over the production of song, but instead has a cumulative effect on song maintenance. The recrystallized songs appeared to remain stable indefinitely. We tracked one bird for a year after his recovery and saw no departures from the baseline song structure.

Our results demonstrate that zebra finches need auditory feedback to maintain their songs in adulthood. This species, which does not modify its song or learn new songs after crystallization, appears to retain a great deal of plasticity in its auditory–vocal control system. This plasticity is sufficient to produce modifications in both

the temporal pattern of song organization and the spectral structure of individual syllables. This finding is not consistent with the classical depiction of song development in which a dynamic learning period in youth ends in a static maintenance period in adulthood. Thus, the distinction between age-limited and open-ended learners may not be as sharp as these names would indicate.

Methods

Birds were housed in custom-designed plexiglass cages and were paired with a female who lived in a separate partition of the cage. The playback speaker's output was calibrated to be approximately the same as the sound level of the bird's vocalizations (80–90 dB SPL in the bird's ear). At the start of the baseline song recordings, the birds ranged from 130 to 300 days old (mean age was 200 days; zebra finches reach adulthood at 90 days). The singing rates of the different birds ranged from tens to thousands of motifs per night. However, there was no apparent correlation of age or singing volume with the magnitude of the effects we observed. For the two birds who did not sing frequently, we collected no baseline data during the feedback period.

Adaptive feedback perturbation. Microphone data were sampled at 40 kHz, after being low-pass filtered (10 kHz cutoff, 7-pole anti-aliasing filter). As is shown in Fig. 1, the computer continuously acquired data from the microphone in segments of 100 ms. Each of these 'bins' of data was passed through a software-based infinite-impulse response (IIR) filter (the 'trigger filter') that was used to detect song vocalizations while avoiding noise artifacts (such as pecking, wing flaps and low amplitude calls). The playback of sound occurred when the output of the trigger filter exceeded a root mean square threshold. The artificial feedback was thus produced with a 100-ms delay, and coincided with a silent interval in the bird's song or with the following syllable, depending on the timing of the song with respect to the bin borders created by the computer. The alternation between recording vocalizations and playing sounds prevented a positive feedback loop from developing. The playback stimulus was constructed from the 100 ms of sound that had caused the trigger event, and was narrowband-filtered before being played back to the bird. A narrowband filter was chosen to make the playback sounds difficult to localize. For one bird a wideband IIR filter was used. No significant differences in results were observed between this bird and the two narrowband birds. Delayed feedback was used instead of other interfering stimuli such as white noise so as to replicate the structure of syllables that these birds normally hear. The effects

of white noise will be examined in future work.

The birds exposed to the adaptive protocol showed changes in their songs similar to those of deafened birds. Because deafening can be caused by prolonged exposure to excessively loud sounds⁷, we used a behavioural test to demonstrate that the conditions used in our experiment did not cause deafness. Zebra finches respond to sounds by vocalizing. At the beginning and the end of the feedback period, we examined differences in the calling probability of two of the adaptive protocol birds to the playback of quiet sounds versus no sounds. The sounds were a variety of natural stimuli including conspecific calls and sounds. Both birds produced significantly more calls during the presentation of quiet sounds ($P < 0.001$, generalized likelihood-ratio test for Bernoulli random variables), which indicates that the birds could hear the sounds.

Syllable-triggered perturbation. Recognition of the target syllable was achieved by using a series of IIR filters in conjunction with each other to perform a logical operation (for example, power in band X and not in band Y). The triggering was based on a small segment of the time-varying spectrum of the target syllable which was unique to that syllable. An original copy of the crystallized trigger syllable was used as the playback stimulus for the duration of the experiment. Typical zebra finch syllables are 80–150 ms in length. The triggering resolution was 50 ms, which was short enough to ensure that the feedback always overlapped the trigger syllable itself (and, partially, the following syllable). The other syllables of the song received no feedback.

Spectral analysis. We calculated the time-frequency spectrogram for each song with a sliding window (5–8 ms) in which each time point consisted of the direct multitaper estimate of the power spectrum (with a time-bandwidth product NW of 3 or 4) (ref. 13). The data shown in Fig. 2 were analysed in this manner. A harmonic analysis was then used to determine the location and magnitude of the jumps in the discrete spectrum of each syllable by calculating the F-spectrum¹⁴ of each of the multitapered spectral estimates. This analysis revealed additional statistically significant harmonic frequencies in the target syllable after the feedback was presented to the bird ($P < 0.01$) (Fig. 3).

Syllable classification. For each syllable, we extracted the length and a number of time-varying parameters (envelope, peak frequency, pitch, goodness-of-pitch and Wiener entropy¹³) based on the spectral analysis described above. A modified K-means clustering algorithm¹⁵ was then used to partition the syllables produced on a given day into subsets. These subsets were labelled by the experimenter (syllable A, syllable B and so on; labelling was done blind to the day on which the data were acquired). For each day of data, approximately 1,000 syllables were analysed. The standard deviations, which are shown as error bars in Fig. 4, were obtained by bootstrapping the probability estimates from the data¹⁶.

Received 23 February; accepted 12 April 1999.

- Konishi, M. The role of auditory feedback in the control of vocalization in the White-crowned sparrow. *Z. Tierpsychol.* **22**, 770–783 (1965).
- Nottebohm, F. Auditory experience and song development in the chaffinch, *Fringilla coelebs*. *Ibis* **110**, 549–569 (1968).
- Sossinka, R. & Bohner, J. Song types in the zebra finch. *Z. Tierpsychol.* **53**, 123–132 (1980).
- Marler, P. & Peters, S. A sensitive period for song acquisition in the song sparrow, *Melospiza melodia*—a case of age limited learning. *Ethology* **76**, 89–100 (1987).
- Nordeen, K. & Nordeen, E. Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches. *Behav. Neural Biol.* **57**, 58–66 (1992).
- Okanoya, K. & Yamaguchi, A. Adult bengalese finches require real-time auditory feedback to produce normal song syntax. *J. Neurobiol.* **33**, 343–356 (1997).
- Wooley, S. & Rubel, E. Bengalese finches *Lonchura striata domestica* depend on auditory feedback for the maintenance of song. *J. Neurosci.* **17**, 6380–6390 (1997).
- Suthers, R. Contributions to birdsong from the left and right sides of the intact syrinx. *Nature* **347**, 473–477 (1990).
- Fee, M. S., Shraiman, B., Pesaran, B. & Mitra, P. P. The role of nonlinear dynamics of the syrinx in birdsong production. *Nature* **395**, 67–71 (1998).
- Houde, J. F. & Jordan, M. I. Sensorimotor adaptation in speech production. *Science* **279**, 1213–1216 (1998).
- Feller, W. *An Introduction to Probability Theory and its Applications* (Wiley, New York, 1968).
- Cover, T. M. & Thomas, J. A. *Elements of Information Theory* (Wiley, New York, 1991).
- Ho, C. E., Pesaran, B., Fee, M. S. & Mitra, P. P. Characterization of the structure and variability of zebra finch song elements. *Proc. 5th Joint Symp. on Neural Computation* 76–83 (1998).
- Thomson, D. J. Spectrum estimation and harmonic analysis. *Proc. IEEE* **70**, 1055–1096 (1982).
- Selmin, S. Z. & Ismail, M. A. K-means-type algorithms: a generalized convergence theorem and characterization of local optimality. *IEEE Trans. Pattern Anal. Mach. Intell.* **1**, 81–87 (1984).
- Bradley, E. *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993).

Acknowledgements. We thank B. Pesaran and M. Sahani for discussions on data analysis, and R. Egnor, M. Fee, G. Laurent and M. Schmidt for comments on the manuscript. This work was supported by a grant from the NIMH.

Correspondence and requests for materials should be addressed to A.L. (e-mail: leonardo@cns.caltech.edu).

Neuronal correlates of parametric working memory in the prefrontal cortex

Ranulfo Romo, Carlos D. Brody, Adrián Hernández & Luis Lemus

Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, México D.F. 04510, México

Humans and monkeys have similar abilities to discriminate the difference in frequency between two mechanical vibrations applied sequentially to the fingertips^{1–3}. A key component of this sensory task is that the second stimulus is compared with the trace left by the first (base) stimulus, which must involve working memory. Where and how is this trace held in the brain? This question was investigated by recording from single neurons in the prefrontal cortex of monkeys while they performed the somatosensory discrimination task. Here we describe neurons in the inferior convexity of the prefrontal cortex whose discharge rates varied, during the delay period between the two stimuli, as a monotonic function of the base stimulus frequency. We describe this as ‘monotonic stimulus encoding’, and we suggest that the result may generalize: monotonic stimulus encoding may be the

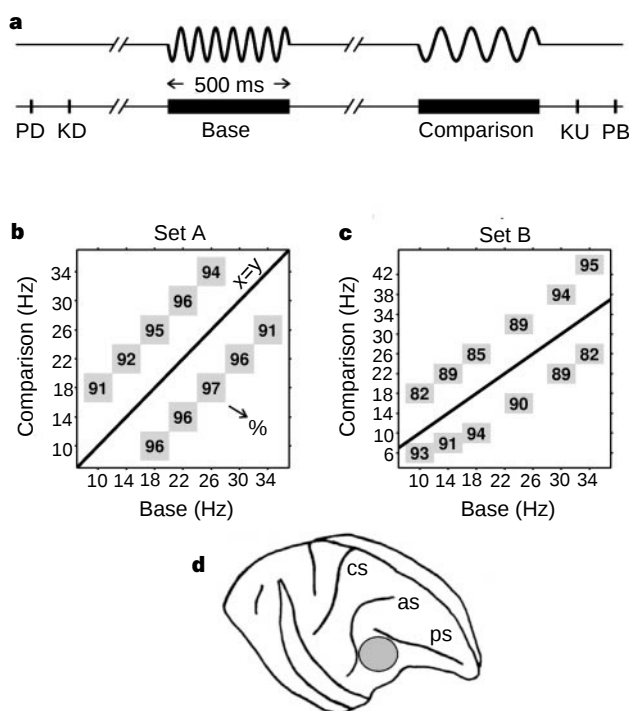


Figure 1 Discrimination task. **a**, Sequence of events during discrimination trials. The mechanical probe is lowered, indenting the glabrous skin of one digit of the hand (PD); the monkey places his free hand on an immovable key (KD); the probe oscillates vertically, at the base frequency; after a delay, a second mechanical vibration is delivered at the comparison frequency; the monkey releases the key (KU) and presses one of two push-buttons (PB) to indicate whether the comparison frequency was higher or lower than the base. **b**, **c**, Stimulus sets used during recordings. Each grey box indicates a base frequency/comparison frequency stimulus pair used; the number inside the box indicates overall per cent correct trials for that base/comparison pair. **d**, Location of recording sites that gave somatosensory working-memory-related responses; CS, central sulcus; AS, arcuate sulcus; PS, principal sulcus.

basic representation of one-dimensional sensory stimulus quantities in working memory. Thus we predict that other behavioural tasks that require ordinal comparisons between scalar analogue stimuli would give rise to monotonic responses similar to those reported here.

During the vibrotactile discrimination task^{1–3} shown in Fig. 1a, animals pay attention to the frequency of the first (base) stimulus, store some trace of it during the delay between the two stimuli and compare the stored trace to the frequency of the second (comparison) stimulus. This task is a parametric working memory task, in the sense that it requires the memorization of a continuous parameter—stimulus frequency (a scalar analogue value induced through a somatic sensation); this is followed, after a delay, by an ordinal comparison of the remembered value to a second scalar analogue—value (also induced through a somatic sensation). Four monkeys (*Macaca mulatta*) were trained to perform the task up to their psychophysical thresholds (on the order of a 2–4 Hz difference between base and comparison for the range of stimulus frequencies used)². After training, neurophysiological recordings were made while the task was performed. Single neuronal activity was recorded through seven independently moveable microelectrodes⁴, separated from each other by 500 μm and inserted in parallel into the prefrontal cortex (PFC). To avoid variations in task difficulty, the two standard stimulus sets used during recording (Fig. 1b, c) had large differences between base and comparison compared by the monkeys' psychophysical threshold. All stimuli were kept within the frequency range which, in humans, gives rise to the percept known as 'flutter'. Vibration frequencies outside this range give rise to qualitatively different percepts and are transduced, in both humans

and monkeys, by cutaneous receptors different from those that transduce flutter⁵.

In pilot experiments with the first monkey, we explored different portions of the left PFC around the principal sulcus. We found neurons which responded during the task only in the inferior convexity of the PFC (Fig. 1d). This region contains neurons that code stimulus identity during visual working memory tasks⁶. Following the pilot experiments, we focused recordings from all animals on this region of the PFC. Although many neurons responded in a task-related manner during the base stimulus or during or after the comparison stimulus, here we focus exclusively on delay period activity.

With the first three monkeys, we used stimulus set A, shown in Fig. 1b. Out of 493 neurons judged during recording to have task-related responses, 318 neurons (65%) were found, based on off-line statistical tests, to have discharge rates during the delay period that varied as a monotonic function of the frequency of the base stimulus. Some of them discharged most weakly after stimulation with the lowest base frequency, and increased their firing rates steadily for increasing base frequencies ($n = 160$, 50%). We refer to these as 'positive monotonic' (Fig. 2a, c, e). Others had discharge rates that varied in the opposite direction ($n = 141$, 44%). We refer to these as 'negative monotonic' (Fig. 2b, d, f). A minority of the neurons ($n = 17$, 5%) changed the sign of their signal during the delay period, being positive monotonic at one time and negative monotonic at a different time. Forty per cent of the time the neuronal firing rate was well described as a linear function of the base stimulus frequency; 21% of the time it was best described as a 'soft' sigmoid function, and 39% of the time it was best described as

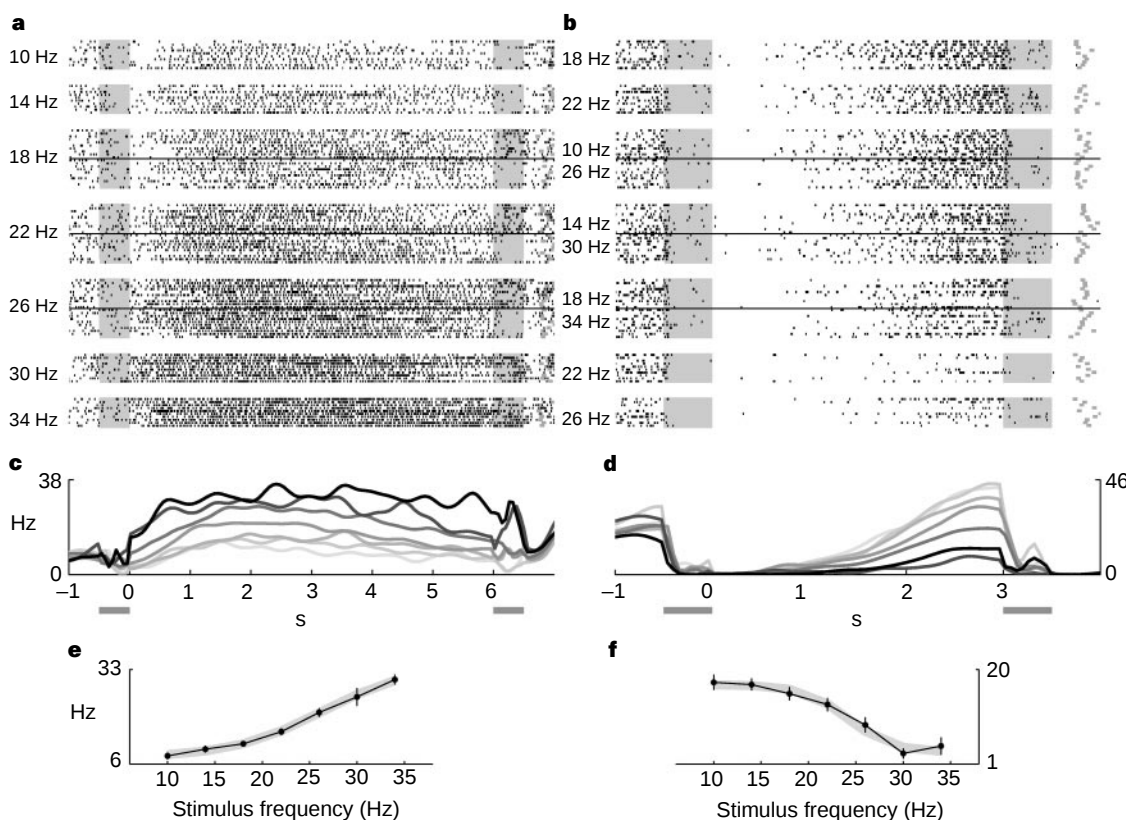


Figure 2 Monotonic responses during the delay period. **a, c, e**, Positive; **b, d, f**, Negative. **a, b**, Rasters. Each row of ticks represents a trial, and each tick represents an action potential. Trials were delivered in a random order, but have been sorted here into blocks of equal base stimulus frequency, indicated on the left. Trials have been further sorted into groups of equal comparison frequency (indicated in the centre), separated by a horizontal black line. Grey boxes (and short horizontal grey lines in **c** and **d**) indicate base and comparison stimulus

periods; thick grey ticks after the comparison stimulus indicate the beginning of the motor response (KU; see Fig. 1a). Time axes for **a** and **b** are shown in **c** and **d**, respectively. **c, d**, Time-dependent spike densities for each base frequency stimulus condition. Grey level indicates the base frequency: the lightest grey line corresponds to 10 Hz and the darkest grey line corresponds to 34 Hz. **e, f**, Mean firing rates, averaged across the entire delay period. Small vertical lines are $\pm$ s.e.m. Thick grey lines are 'soft' sigmoid fits to the data (see Methods).

a 'hard' sigmoid function (see Methods). Thus, the base stimulus frequency (a scalar analogue value) appeared to be encoded directly in the neurons' firing rate (also a scalar analogue value), most often in a smoothly graded fashion.

Response characteristics during the delay period were not static^{7,8} (Fig. 2b). Most of the neurons ($n = 260$, 82%) were studied using a fixed (3-s) delay period. With a fixed delay period, the monkey can anticipate the timing of the second stimulus, and neuronal activity often reflected this fact (Figs 2b, 3e, f). For each neuron, we determined the times during the delay period in which its firing rate encoded significant monotonic signal about the base stimulus (see Methods). Most neurons could be described as falling into three main groups, illustrated in Fig. 3. 'Early' neurons (Fig. 3a, b) carried a signal about the base stimulus frequency during the first second, but not during the last second, of the delay period ($n = 89$, 34%). 'Persistent' neurons (Figs 3c, d, 2a) carried a signal about the base stimulus during all three seconds of the delay period ($n = 60$, 23%). Finally, 'late' neurons (Figs 3e, f, 2b) carried a signal about the base stimulus during the last second, but not during the first second, of the delay period ($n = 85$, 33%). A smaller number of neurons ($n = 26$, 10%) could not be assigned to any of the three main groups. Figure 3g shows the total number of neurons that were determined to carry significant monotonic signal, as a function of time. A decrease following the beginning of the delay period is followed by a marked rise towards its end, reflecting the large number of 'late' neurons. During recording, no attempt was made to sample one or other of the timing groups preferentially. The shape shown in Fig. 3g was found after the end of the recording sessions.

Forty-three of the neurons studied with a 3-s delay period were observed for long enough to subsequently study them with a complete set of 6-s delay period trials. The results showed that 'late' neurons behave similarly to those described in a visuomotor working memory task³: the time at which 'late' neurons carry the strongest signal about the base stimulus is not locked to the beginning of the delay period, but shifts in proportion to the length of the delay period, thus anticipating its end (R.R. *et al.*, unpublished results).

Were PFC neurons of the inferior convexity encoding information about the base stimulus itself, or were they encoding an anticipated motor act? Let P denote the probability that the correct motor act in a given trial will be to press the button indicating that the comparison frequency was lower than the base frequency. For stimulus set A (Fig. 1b), P is a monotonic function of the base frequency, ranging from 0 (base frequencies 10 and 14 Hz) to 1 (base frequencies 30 and 34 Hz). Therefore, monotonic neurons recorded from while using stimulus set A may be encoding either the remembered base stimulus or the probability of an anticipated motor act. To distinguish between these two possibilities, we recorded from the PFC of a fourth monkey, also trained in the vibrotactile task, while it responded to stimulus set B (Fig. 1c). This is a stimulus set specifically designed so that P was constant at 0.5 for each base frequency. Thus, the base stimulus frequency carried no information about the required motor act. We nevertheless found delay period responses very similar to those obtained using stimulus set A: out of 123 neurons judged during recording to show task-related responses, 85 neurons (69%) had firing rates during the delay period that were significantly monotonic functions of the base stimulus frequency. Fifty-nine of these (69%) were positive monotonic, and 25 (29%) were negative monotonic (one neuron changed sign during the delay period). Thirty-eight per cent of the neurons were classified as 'early', 35% as 'persistent' and 19% as 'late'. Sixty-nine per cent of the time the responses were smooth functions (linear or soft sigmoid) of the base frequency. The most parsimonious hypothesis, covering all the monotonic responses found in all monkeys using several different stimulus sets (including sets A and B), is that monotonic neurons were encoding the memorized base

stimulus frequency.

The conclusion that some monotonic neurons can encode the base stimulus frequency applied to all monotonic neuron types, including 'late' neurons. These share the anticipatory timing of their base-stimulus dependence (Fig. 3e, f) with other anticipatory neurons, found in various regions of the brain^{8,10} and in the PFC^{9,11,13} during delayed response tasks. Such neurons can usually be interpreted as anticipatory motor neurons^{7,10,14,15}. However, the use of stimulus set B reveals that, entirely separately from encoding the timing^{7,16} or identity⁸⁻¹³ of an impending motor act, the firing of some anticipatory neurons in the PFC can also encode the memorized stimulus itself. This challenges their interpretation as neurons involved in motor preparation. Are 'late' neurons in the PFC important for non-motor computations which use the memory trace? What mechanisms modulate their late acquisition of the base stimulus signal? We do not expect these questions to be unique to the vibrotactile discrimination task.

The smooth monotonic encoding found in many neurons is consistent with the existence of a parametric, rather than categorical, representation of the memorized stimulus during the working memory task¹⁷. We suggest that monotonic encodings may be the

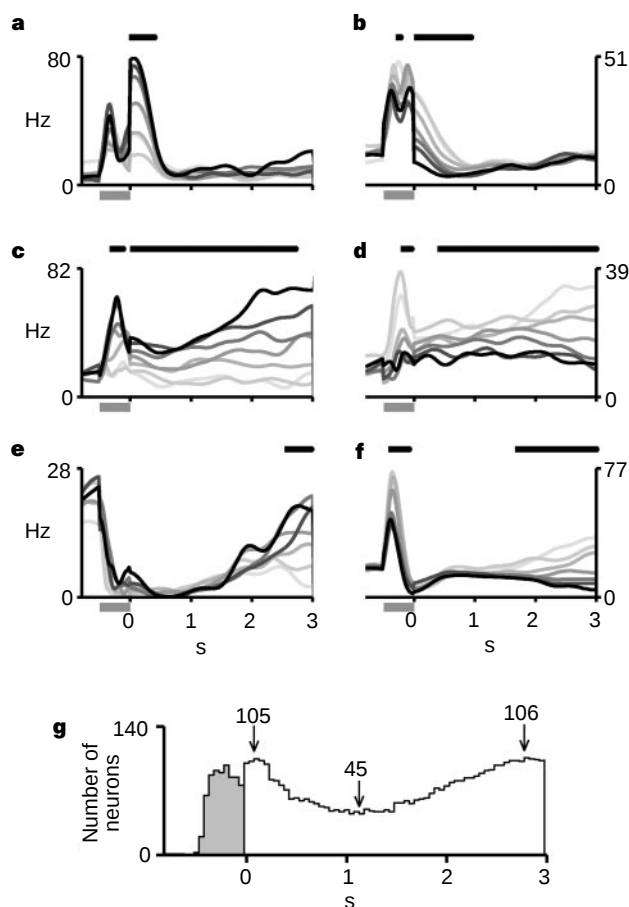


Figure 3 Neuronal response types. **a-f**, Single-neuron spike density functions from six different neurons, with conventions as in Fig. 2c, d. The dark bars above each plot indicate times during which the neuron's firing rate carried significant ($P < 0.01$) monotonic signal about the base stimulus. **a, c, e**, Positive monotonic neurons. **b, d, f**, Negative monotonic neurons. **g**, Total number of recorded neurons (during fixed 3-s delay period runs) carrying a significant signal about the base stimulus, as a function of time relative to the beginning of the delay period. Individual neurons may participate in more than one bin. The base stimulus period is shaded grey, and the maximum, minimum, and maximum number of neurons, during the first, middle, and last seconds of the delay period, respectively, are indicated with arrows.

basic representation of sensory magnitude continua during working memory. For the task used here, the monotonic encoding may be derived from inputs from cortical somatosensory areas^{18,19}: in recordings from the primary and secondary somatosensory cortices during the same task²⁰, we found neurons in the second somatosensory cortex that responded in a manner similar to 'early' neurons, encoding the base stimulus frequency monotonically in their firing rates, and carrying signal into the early part of the delay period. Although many neurons in the primary somatosensory cortex responded during the stimulus periods, none did so during the delay period^{1,20}, in contrast to results obtained during a haptic (touch) task²¹. As in other working memory tasks^{7,8}, the firing rate of many neurons during the delay period was not only a function of the memorized stimulus but also a function of time (Figs 2b, 3). For such neurons, the stimulus cannot be identified on the basis of firing rate alone: decoding the information carried by these neurons requires knowledge of the time that has elapsed since the stimulus, or comparison between the firing rates of neurons with similar time-courses but different stimulus dependencies. Precisely how (or even whether) time-dependent neurons are in fact decoded, and their precise role in working memory are unknown. Our results contribute two items to the current debate regarding the functional anatomy of the PFC^{6,22,23}. First, they support the view that working memory responses for tasks that do not include a spatial component are preferentially found in the ventral regions of the lateral PFC⁶; however, we explored the dorsolateral PFC in only one monkey. Second, the results constitute an electrophysiological demonstration that neurons in the PFC can retain working memory information induced by non-visual²⁴ as well as visual^{7,11,25} modalities. Whether this multimodal capacity is subserved by intermingled but modality-specific neurons, or whether it is subserved by individually multimodal neurons, remains to be investigated in paradigms involving more than one sensory modality. □

Methods

Animals were handled according to the institutional standards of the NIH and Society for Neuroscience.

Discrimination task. The discrimination task used here has been described^{2,3}. Briefly, stimuli were delivered to the skin of the distal segments of one digit of the right, restrained hand, via a computer-controlled Chubbuck linear motor stimulator (2-mm round tip). The initial indentation was 500 μ m. Stimulation amplitudes were adjusted to produce equal subjective intensities. During trials, two vibrotactile stimuli were delivered consecutively to the glabrous (hairless) skin, separated by an interstimulus delay period (Fig. 1a), and the animal was rewarded for correct discrimination with a drop of liquid. Discrimination was indicated by pressing one of two push-buttons. Performance was measured with psychophysical techniques¹⁻³.

Recording sessions and sites. Neuronal recordings were obtained with an array of seven independent microelectrodes^{3,4} (2–3 M Ω) inserted into the inferior convexity of the PFC of the left hemisphere of three monkeys and in the right hemisphere in two monkeys. Recordings sites changed from session to session. We used standard histological procedures to construct surface maps of all penetrations in the PFC. This was done first by marking the edges of the small chamber (7 mm diameter) placed above the PFC. Additionally, in the last recording sessions, we made small lesions at different depths. Electrode penetrations were normalized against the arcuate sulcus and principal sulcus. The penetrations that gave the results reported here were located in the inferior convexity of the PFC (Fig. 1d).

Data analysis. A continuous-time data analysis method was used. Single-neuron spike trains were convolved with gaussian kernels to obtain time-dependent spike-density functions for each trial. (Kernels were corrected for edge effects and stimulus period densities were computed separately from densities of other periods; stimulus periods kernel $\sigma = 50$ ms, other periods $\sigma = 150$ ms: thus, the effective timescale of the delay period analysis was ~ 300 ms.) A time-dependent spike-rate mean and a time-dependent spike-rate standard error of the mean (s.e.m.) were computed from the set of density functions for each base stimulus condition. At each point in time, we computed

the best linear fit of mean response as a function of base stimulus frequency; the significance of the slope of the linear fit being different from zero was then computed from the standard errors^{26,27}. We also computed the χ^2 goodness-of-fit probability Q of both a linear (2 degrees of freedom) and a sigmoidal (4 degrees of freedom) fit to the data²⁷. Response times where the linear regression slope was significantly different from zero ($P < 0.01$) and either or both of the linear and sigmoidal fits were acceptable ($Q > 0.05$) were marked as 'significantly monotonic'. These times were further marked as either 'linear' or 'sigmoidal', according to which fit had the higher Q . A 'soft' sigmoid was defined as a sigmoid which was approximately linear over at least half the range of base stimulus frequencies; sigmoids linear over a shorter range of values were called 'hard' sigmoids. Neurons with significantly monotonic times in the delay period were marked as delay period signal-carrying neurons. To evaluate the net effect of our marking methods, we took 318 significantly monotonic neurons and shuffled the base frequency labels among trials of each neuron. Upon re-analysis, only 13 of the shuffled neurons were marked as significantly monotonic. Thus, the net probability of marking a neuron as monotonic by chance was approximately $P = 13/318 \approx 0.04$.

Received 13 January; accepted 6 April 1999.

- Mountcastle, V. B., Steinmetz, M. A. & Romo, R. Frequency discrimination in the sense of the flutter: psychophysical measurements correlated with postcentral events in behaving monkeys. *J. Neurosci.* **10**, 3032–3044 (1990).
- Hernandez, A., Salinas, E., Garcia, R. & Romo, R. Discrimination in the sense of flutter: new psychophysical measurements in monkeys. *J. Neurosci.* **17**, 6391–6400 (1997).
- Romo, R., Hernandez, A., Zainos, A. & Salinas, E. Somatosensory discrimination based on cortical microstimulation. *Nature* **392**, 387–390 (1998).
- Mountcastle, V. B., Reitboeck, H. J., Poggio, G. F. & Steinmetz, M. A. Adaptation of the Reitboeck method of multiple microelectrode recording to the neocortex of the waking monkey. *J. Neurosci. Methods* **36**, 77–84 (1991).
- Talbot, W. H., Darian-Smith, I., Kornhuber, H. H. & Mountcastle, V. B. The sense of flutter-vibration: comparison of the human capacity with response patterns of mechanoreceptive afferents from the monkey hand. *J. Neurophysiol.* **31**, 301–334 (1968).
- Wilson, F. A. W., O'Scalaidhe, S. P. & Goldman-Rakic, P. S. Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* **260**, 1955–1958 (1993).
- Fuster, J. M. *The Prefrontal Cortex* 2nd edn (Raven, New York, 1989).
- Chafee, M. V. & Goldman-Rakic, P. S. Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working-memory task. *J. Neurophysiol.* **79**, 2919–2940 (1998).
- Kojima, S. & Goldman-Rakic, P. S. Delay-related activity of prefrontal neurons in rhesus-monkeys performing delayed-response. *Brain Res.* **248**, 43–49 (1982).
- Bruce, C. J. & Goldberg, M. E. Primate frontal eye fields. I. Single neurons discharging before saccades. *J. Neurophysiol.* **53**, 603–635 (1985).
- Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* **61**, 331–349 (1989).
- Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. *J. Neurophysiol.* **63**, 814–831 (1990).
- Hoshi, E., Shima, K. & Tanji, J. Task-dependent selectivity of movement-related neuronal activity in the primate prefrontal cortex. *J. Neurophysiol.* **80**, 3392–3397 (1998).
- Quintana, J. & Fuster, J. M. Mnemonic and predictive functions of cortical neurons in a memory task. *Neuroreport* **3**, 721–724 (1992).
- Asaad, W. A., Rainer, G. & Miller, E. K. Neural activity in the primate prefrontal cortex during associative learning. *Neuron* **21**, 1399–1407 (1989).
- Schultz, W. & Romo, R. Role of primate basal ganglia and frontal cortex in the internal generation of movements. I. Preparatory activity in the anterior striatum. *Exp. Brain Res.* **91**, 363–384 (1992).
- Seung, H. S. How the brain keeps the eyes still. *Proc. Natl Acad. Sci. USA* **93**, 13339–13344 (1996).
- Pandya, D. N. & Barnes, C. L. In *The Frontal Lobes Revisited* (ed. Perecman, E.) 41–72 (IRBN, New York, 1987).
- Preuss, T. M. & Goldman-Rakic, P. S. Connections of the ventral granular frontal cortex of macaques with perisylvian premotor and somatosensory areas: anatomical evidence for somatic representation in primate frontal association cortex. *J. Comp. Neurol.* **282**, 293–316 (1989).
- Salinas, E., Hernández, A., Zainos, A., Lemus, L. & Romo, R. Cortical recoding of sensory stimuli during somatosensory discrimination. *Soc. Neurosci. Abstr.* **24**, 1126 (1998).
- Zhou, Y. & Fuster, J. M. Mnemonic neuronal activity in somatosensory cortex. *Proc. Natl Acad. Sci. USA* **93**, 10533–10537 (1996).
- Rao, S. C., Rainer, G. & Miller, E. K. Integration of what and where in the primate prefrontal cortex. *Science* **276**, 821–824 (1997).
- Rainer, G., Asaad, W. F. & Miller, E. K. Memory fields of neurons in the primate prefrontal cortex. *Proc. Natl Acad. Sci. USA* **95**, 15008–15013 (1998).
- Bodner, M., Kroger, J. & Fuster, J. M. Auditory memory cells in dorsolateral prefrontal cortex. *Neuroreport* **96**, 1905–1908 (1996).
- Miller, E. K., Erickson, C. A. & Desimone, R. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* **16**, 5154–5167 (1996).
- Ross, S. M. *Introduction to Probability for Scientists and Engineers* (Wiley, New York, 1987).
- Press, W., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in C* 2nd edn (Cambridge Univ. Press, Cambridge, U.K., 1992).

Acknowledgements. The research of R.R. was partially supported by an International Research Scholars Award from the Howard Hughes Medical Institute and grants from DGAPA-UNAM, CONA-CyT and Fundación Miguel Alemán A.C. We thank W. Newsome, M. Shadlen, their research groups, and V. Mountcastle and E. Salinas for extensive comments and discussion; A. Zainos for help during the experiments; and F. Jandete and S. Méndez for technical assistance. R.R. conceived the experiment, and together with A.H. and L.L. carried it out; C.D.B. designed and carried out the data analysis; R.R. and C.D.B. co-wrote the paper.

Correspondence and requests for materials should be addressed to R.R. (e-mail: rromo@ifisil.unam.mx).

Developmental basis of limblessness and axial patterning in snakes

Martin J. Cohn*† & Cheryll Tickle†‡

* Division of Zoology, School of Animal and Microbial Sciences, University of Reading, Whiteknights, Reading RG6 6AJ, UK

‡ Department of Anatomy and Physiology, Wellcome Trust Building, University of Dundee, MSI/WTB Complex, Dow Street, Dundee DD1 5EH, UK

† Department of Anatomy and Developmental Biology, University College London, Medawar Building, Gower Street, London WC1E 6BT, UK

The evolution of snakes involved major changes in vertebrate body plan organization, but the developmental basis of those changes is unknown. The python axial skeleton consists of hundreds of similar vertebrae, forelimbs are absent and hindlimbs are severely reduced. Combined limb loss and trunk elongation is found in many vertebrate taxa¹, suggesting that these changes may be linked by a common developmental mechanism. Here we show that Hox gene expression domains

are expanded along the body axis in python embryos, and that this can account for both the absence of forelimbs and the expansion of thoracic identity in the axial skeleton. Hindlimb buds are initiated, but apical-ridge and polarizing-region signalling pathways that are normally required for limb development are not activated. Leg bud outgrowth and signalling by Sonic hedgehog in pythons can be rescued by application of fibroblast growth factor or by recombination with chick apical ridge. The failure to activate these signalling pathways during normal python development may also stem from changes in Hox gene expression that occurred early in snake evolution.

Limblessness has evolved many times in vertebrate evolution, and is often accompanied by elongation and loss of regional differentiation in the axial skeleton¹. Snakes evolved from tetrapod lizards and are closely related to mosasaurs, which are Cretaceous marine lizards that had complete forelimbs and hindlimbs and a clearly regionalized axial skeleton^{1,2}. Pythons have over 300 vertebrae (Fig. 1a), with ribs on every vertebra anterior to the hindlimbs, except for the atlas (Fig. 1a, b, e). The anterior vertebrae have both ribs (a thoracic feature) and ventral hypopophyses (generally a cervical feature) (Fig. 1b), suggesting that information encoding thoracic identity may have extended into the cervical region and partially transformed these segments. Thus, the entire trunk resembles an elongated thorax (Fig. 1a). There is no morphological evidence of forelimb development. Functional rudimentary hind-

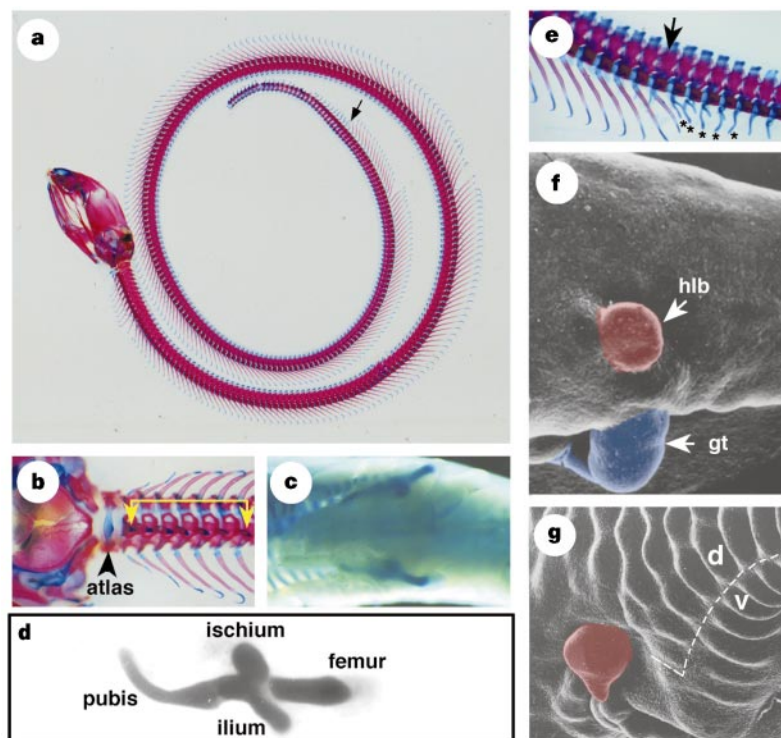


Figure 1 Morphology of python axial and appendicular skeleton. Anterior is to the left. **a**, **b**, **e**, Alcian blue and Alizarin red stained skeletal preparation of python embryo at 24 days of incubation. Arrows mark position of hindlimb rudiments, which have been removed to improve visibility of vertebrae. **a**, Lateral view of complete skeleton. Note homogeneity of vertebrae anterior to arrow. **b**, High-power ventral view of anterior axial skeleton and base of skull. Atlas lacks ribs (arrowhead) and hypopophyses (bracketed), which extend ventrally from 64 vertebral bodies posterior to atlas. **c**, Ventral view of python embryo at 14 days of incubation stained with Alcian green to reveal skeletal pattern. The pelvis is visible within the body wall and short femora protrude from the body wall. **d**, Skeletal preparation of hindlimb and associated pelvis dissected from embryo shown in **c**. Femur and all three elements of pelvic girdle are present (pubis, ilium and

ischium). **e**, Skeletal preparation showing cloacal region of python embryo at 24 days of incubation. Arrow indicates position of the hindlimb (removed) relative to axial skeleton. Hindlimb position corresponds to a transitional vertebra with intermediate morphology (arrow), separating vertebrae with large, movable ribs (left) from vertebrae with lymphapophyses in cloacal region (right, with asterisks). **f**, Scanning electron micrograph of left limb bud and trunk of python embryo at 4 days of incubation. Hindlimb bud (hlb; red) lies dorsal to the paired genital tubercles (gt; blue) which develop on the left and right margins of the cloaca. **g**, Scanning electron micrograph of python embryo at 24 days of incubation showing left hindlimb (red); interface between morphologically distinct dorsal scales (d) and ventral belly scales (v), marked by dashed line.

limbs, consisting of a pelvic girdle and truncated femur (Fig. 1c, d), develop at the junction between rib-bearing vertebrae and lymphopophysis-bearing vertebrae, at cloacal level (Fig. 1e, f).

Hox genes specify the axial pattern during embryonic development. In animals with different numbers of vertebrae, Hox expression domains in the paraxial mesoderm (which gives rise to vertebrae) correlate with vertebral identity rather than number^{3,4}. Hox genes also appear to be involved in the regionalization of the lateral plate mesoderm into forelimb, flank and hindlimb, to specify limb position^{5,6}. To determine whether changes in Hox gene expression in the paraxial and lateral plate mesoderm underlie the morphological transformations seen in the python trunk, we examined the distribution of three Hox proteins, HOXC6, HOXC8 and HOXB5. HOXC6 and HOXC8 are associated with the development of thoracic vertebrae in other tetrapods^{7,8} (Fig. 2a), whereas HOXB5 is expressed up to the first cervical vertebra (the atlas)^{6,9}. In the lateral plate mesoderm of tetrapods and fish, the anterior expression boundaries of all three genes occur at the forelimb/pectoral fin level, where they are involved in specifying forelimb position and shoulder development^{6,10,11}. Python eggs are laid about eight weeks after fertilization, by which time the embryos have developed somites (derived from paraxial mesoderm) and hindlimb buds. In these embryos, HOXC6, HOXC8 and HOXB5 are expressed in somites throughout the entire trunk, extending from the cloacal/hindlimb level to the most anterior somite (Fig. 2b–g). We detected a sharp posterior boundary of HOXC8 expression at the level of the hindlimbs (Fig. 2b), which coincides with the last thoracic vertebra in older animals (compare Fig. 2b with Fig. 1a and e). HOXC8 and

HOXB5 are present throughout the python lateral plate mesoderm, with expression terminating at the very anterior limit of the trunk (Fig. 2c, f). Thus, the entire vertebral column anterior to the cloaca exhibits patterns of Hox gene expression consistent with thoracic identity, and we were unable to detect restricted Hox expression patterns in the lateral plate mesoderm associated with forelimb position in other tetrapods (Fig. 2h). Expansion of these Hox gene expression domains in both paraxial and lateral plate mesoderm may be the mechanism which transformed the entire snake trunk towards a thoracic/flank identity and led directly to the absence of forelimb development during snake evolution.

The specification of hindlimb position and initiation of budding appears to be normal in python embryos. The outgrowth of vertebrate limb buds depends on the apical ectodermal ridge, a thickened epithelium rimming the distal edge of the limb buds (Fig. 3b)¹². Although direct-developing frogs undergo normal limb development without forming a distinctive apical ridge, the distal limb bud ectoderm nevertheless expresses genes associated with ridge function¹³. In python embryos between 1 and 5 days of incubation, no apical ridge could be detected either in histological sections or by scanning electron microscopy (Fig. 3a), and products of the *Distal-less* (*Dlx*), *Fgf2* and *Msx* genes, which normally characterize the apical ridge^{14–16}, were not detected in limb ectoderm (Fig. 3c–h). High levels of expression of these genes were detected in python limb mesenchyme and/or other developing organs (for example, kidneys, tooth buds and scale buds), demonstrating that the antibodies we used can recognize the reptilian gene products. The absence of apical ectodermal ridge and lack of

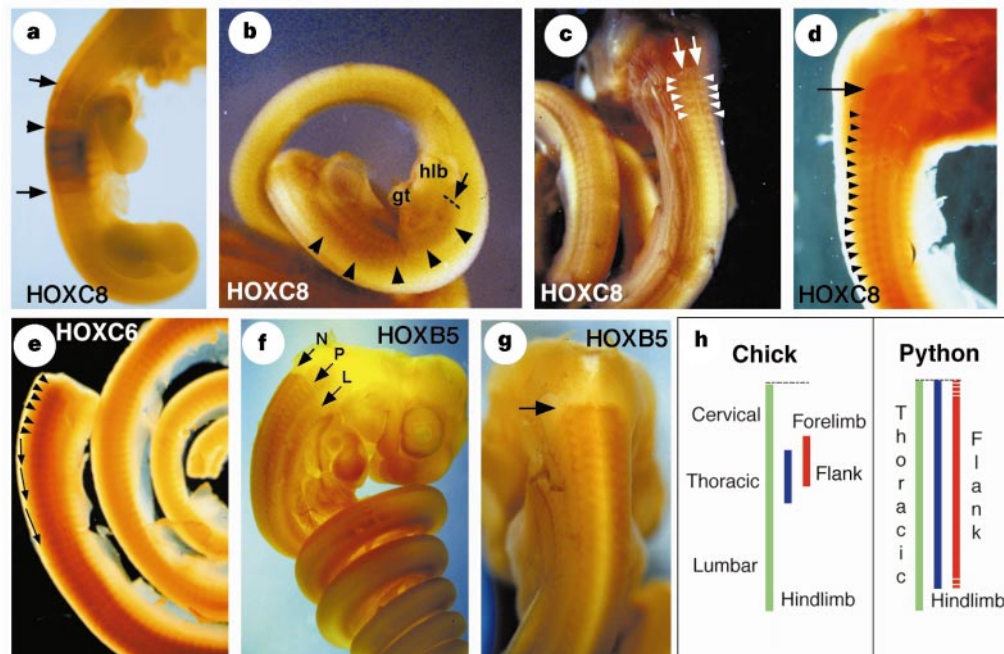


Figure 2 Whole-mount antibody staining showing Hox gene expression in python and chick embryos. Anterior is at the top in **a** and **c–h**, and at bottom in **b**. Positive signal is indicated by darkly stained cells. **a**, HOXC8 expression is restricted to thoracic region in chick embryo at stage 25. Arrow, anterior and posterior boundaries; arrowhead, boundary between strong and weak expression. **b–d**, Expression of HOXC8 throughout the trunk of python embryos at one day incubation. **b**, HOXC8 expression (arrowheads) extends posteriorly to level of hindlimb bud (hlb). Dashed line marks the sharp posterior expression boundary. **c, d**, Anterior HOXC8 expression in cleared (**d**) and uncleared (**c**) python embryos. Segmental pattern of expression (arrowheads) is detected in prevertebrae up to the anterior limit of the axial skeleton (arrows). **e**, HOXC6 expression in a python embryo. Segmental expression (arrowheads) is detected throughout the entire

trunk posterior to the head, which has been removed, in a pattern similar to HOXC8. **f, g**, HOXB5 expression detected throughout trunk of python embryo, with sharp anterior boundary of expression in neural tube at the junction of the spinal cord and hindbrain (arrow in **g**). **f**, Anterior boundaries of expression domains (arrows) in neural tube (N), paraxial (P) and lateral plate (L) mesoderm are in register with one another. **h**, Schematic diagram comparing expression domains of HOXB5 (green), HOXC8 (blue) and HOXC6 (red) in chick and python embryos. Broken line at anterior and posterior extremes of red line indicates lack of certainty about precise limits of HOXC6 expression. Expansion of HOXC8 and HOXC6 domains in python correlates with expansion of thoracic identity in axial skeleton and flank identity in lateral plate mesoderm.

expression of ridge-associated genes could account for the severe hindlimb truncation seen in pythons.

Apical ectodermal ridge signalling is mediated by fibroblast growth factors (FGFs)^{17,18}. Therefore, we tested whether grafting FGF2-loaded beads to python limb bud could sustain outgrowth. The proximodistal length of FGF2-treated limb buds was increased compared with the contralateral buds in two out of five embryos treated. One day after grafting, a 31% increase was seen with one FGF bead and a 9% increase was seen with two FGF beads. Thus, python limb bud outgrowth can be stimulated by FGF. As FGF can also promote proliferation in cultured vestigial limb buds of the slow-worm *Anguis fragilis*¹⁹, a serpentine lizard, our results raise the possibility that the independent evolution of limblessness in different reptilian lineages may have involved similar developmental mechanisms.

In amniote limbs, the apical ridge forms at the boundary of the dorsal and ventral ectodermal compartments²⁰. A dorsoventral

interface between the morphologically distinct dorsal and ventral scales can be seen extending along the entire trunk ectoderm of python embryos, and the paired hindlimb buds develop at this interface (Fig. 1g). The failure of ridge development in chicken *limbless* mutants is due to a lack of dorsoventral polarity in the limb buds²¹. Therefore, we examined python limb buds for expression of two molecules associated with dorsoventral polarity, Engrailed (EN) and LMX1. Just as in normal chick limb buds²² (Fig. 4i), EN was detected in the ventral ectoderm of python limb buds, with a sharp boundary of expression running along the bud apex (Fig. 4h). In addition, LMX1 (ref. 23) was confined to dorsal limb mesenchyme cells in pythons (Fig. 4j). Thus, python leg buds, in contrast to the limb buds of chicken *limbless* mutants, have normal dorsoventral polarity.

Another possible reason for the failure of ridge formation is that mesenchymal changes have occurred in python hindlimb buds. To test the ability of python limb bud mesenchyme to signal to the ectoderm, we transplanted mesenchyme from the posterior of the python limb bud to the anterior chick wing bud, and then monitored expression of chick *Fgf8* to determine the extent of the apical ridge. At 19.5 h after transplantation, *Fgf8* expression was detected in anterior chick limb ectoderm overlying the python graft, whereas expression in the contralateral limb did not extend as far to the anterior (Fig. 4g). Thus, python limb mesenchyme can maintain an apical ectodermal ridge and *Fgf8* expression.

Mesenchymal cells in the polarizing region, located at the posterior margin of vertebrate limb buds, act as a signalling centre and express Sonic hedgehog (SHH), which specifies the anteroposterior pattern of the limb. The polarizing region and apical ridge maintain each other through a positive feedback loop, mediated by SHH and FGF4, which coordinates limb bud outgrowth and patterning^{17,18}. When *Shh* is functionally inactivated in mice, their limbs are truncated²⁴. We therefore examined *Shh* expression in python embryos, which lack an apical ridge. SHH protein is closely associated with *Shh* messenger RNA in the polarizing region in chick and mouse limb buds (Fig. 4b)²⁵. In contrast, no SHH protein could be detected in python hindlimb buds (Fig. 4a). SHH was present, however, in floor plate of the neural tube and in the notochord, both of which are known sites of *Shh* expression in other vertebrate embryos²⁵. Thus, in the absence of an apical ridge, *Shh* is not expressed in python hindlimb buds.

Three different assays, however, showed that python hindlimb bud mesenchyme retains remarkable potential to express SHH and act as a polarizing region. In python posterior mesenchyme cells grafted under the apical ridge of a chick wing bud at stage 20, SHH was expressed within 24 h (Fig. 4e). In a second chick wing bud, expression of the SHH receptor *patched*, which is induced in response to SHH signalling²⁶, was detected in chick limb cells around a python graft 19.5 h after transplantation (Fig. 4f). Finally, a wing bud with a python graft left to develop for seven days contained two additional digits (anterior-to-posterior digit pattern 2-2-2-3-4), compared with the normal pattern (2-3-4) (Fig. 4c). This digit pattern can also be produced by anterior grafts of small numbers of *Shh*-expressing cells²⁷. These data show that python hindlimb mesenchyme is competent to express SHH and send a polarizing signal, and suggest that it fails to do so during python hindlimb development because the apical ridge is absent.

Unexpectedly, anterior mesenchyme grafted from python hindlimb also has polarizing potential and induced the development of an additional digit 2 in a chick wing (Fig. 4d). Thus, polarizing potential appears to exist both anteriorly and posteriorly in python hindlimb buds, rather than being posteriorly restricted as in other vertebrates. Moreover, three grafts of python lateral plate mesenchyme anterior to the hindlimb bud suggested that polarizing potential is also present in the flank. Several lines of evidence indicate that Hox gene-expression domains along the primary body axis define the spatial extent of polarizing potential. For

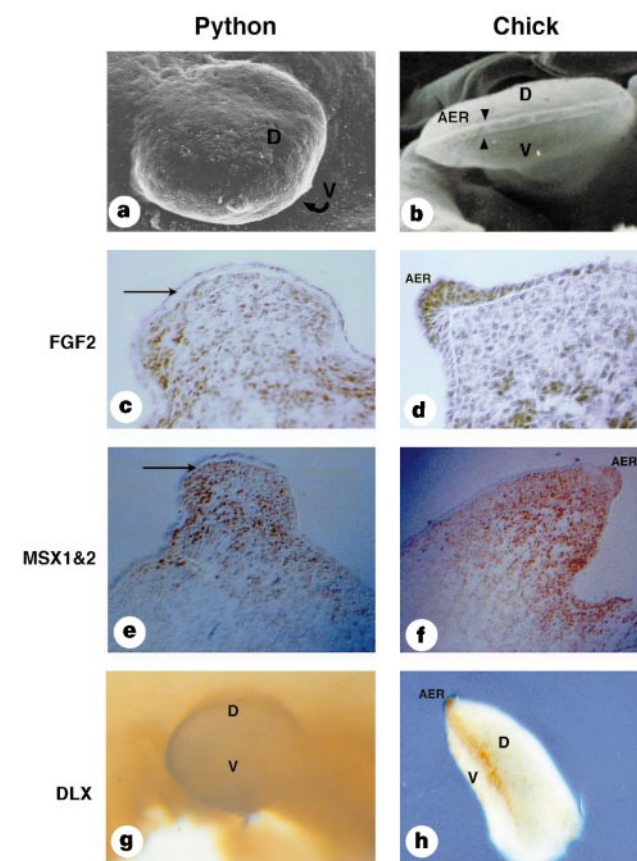


Figure 3 Antibody staining and scanning electron micrographs comparing apical ectoderm in python and chick limb buds. **a, b**, Scanning electron micrographs of a python embryo limb bud at 4 days (**a**) and chick embryo at 5 days of incubation (**b**). Distal tip of python bud lacks apical ectodermal ridge (**a**), in contrast to chick, which has a distinct ridge (AER) at the boundary between dorsal (D) and ventral (V) ectoderm (**b**). Transverse sections of python limb buds at 1 day of incubation (**c, e, g**) and stage 20 chick leg buds (**d, f, h**) stained with antibodies against FGF2, MSX, and DLX. **c**, FGF2 expression is present in python limb bud mesenchyme but could not be detected in the ectoderm (arrow). **d**, FGF2 expression is present in chick limb bud mesenchyme and ectoderm, with strong expression in apical ectodermal ridge. **e**, MSX expression is present in python limb bud mesenchyme in a proximal-to-distal gradient, but not in ectoderm (arrow). **f**, MSX expression is present in chick limb bud mesenchyme in a proximal-to-distal gradient, and in the apical ectodermal ridge (AER). **g**, DLX expression could not be detected in the python limb. **h**, DLX expression in the apical ectodermal ridge (AER) of chick leg bud.

example, anterior extension of the *Hoxb8* domain in transgenic mice induces *Shh* expression anteriorly in the forelimb²⁸. Thus, widespread polarizing potential in pythons is consistent with our finding that Hox expression domains are expanded anteroposteriorly. Furthermore, since positional information in the limb ectoderm is determined initially by the underlying mesoderm²⁹, it is possible that changes in mesodermal Hox gene expression in pythons may also have eliminated the ability of the ectoderm to form an apical ridge.

A simple developmental mechanism involving progressive changes in Hox gene expression along the main body axis could link expansion of thoracic identity with acquisition of limblessness

in snake evolution (Fig. 5). The primitive condition for all squamates, including snakes, is possession of complete forelimbs and hindlimbs, and a relatively short, regionalized vertebral column. The most primitive snake known, *Pachyrhachis problematicus*, had complete (or almost complete) polarized hindlimbs, but no forelimbs, and an elongated vertebral column with ribs on almost every segment². Both scolecophidians and booids (which includes pythons) are primitive snakes with severely reduced hindlimbs and a trunk resembling an elongated thorax. Advanced snakes (colubroids) have even more uniformity in the axial skeleton and are completely limbless. Progressive expansion of Hox gene expression domains along the body axis can account for the major

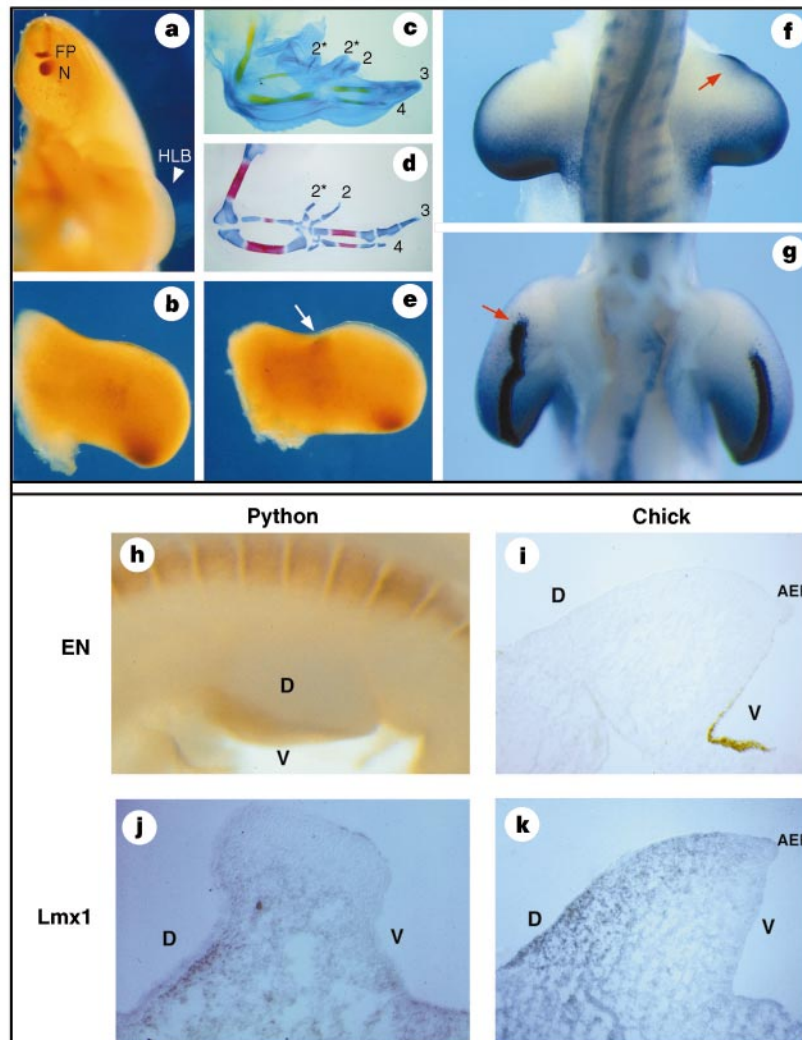


Figure 4 Polarizing activity and dorsoventral polarity in python hindlimb buds. Anterior is at top in **a–g** and at right in **h, a**. Python embryo at 2 days of incubation stained with an antibody against SHH. SHH expression (dark brown) detected in floor plate (FP) and notochord (N), but not in hindlimb bud (HLB). **b**, SHH expression in polarizing region of chick wing bud. **c**, Wing of ten-day chick embryo with duplicated pattern of digits that developed after transplantation of python posterior limb bud mesenchyme to anterior margin of wing bud at stage 20. Two additional digit 2s were specified anterior to the normal set of digits. (asterisk, duplicated digits). **d**, Wing of ten-day chick embryo with duplicated pattern of digits that developed after transplantation of python anterior limb bud mesenchyme to anterior margin of wing bud at stage 20. A single duplicated digit 2 is present anterior to normal digits. **e**, SHH expression in chick wing bud 24 h after python posterior limb mesenchyme was grafted anteriorly under the apical ridge. SHH expression was detected in the graft of python cells (arrow) and in the host

chick polarizing region. **f, g**, Double *in situ* hybridization showing expression of chick *Ptc* and *Fgf8* 24 h after transplantation of python posterior limb mesenchyme to anterior margin of the chick right wing bud. Red arrow in **f** (dorsal view) indicates ectopic expression of chick *Ptc* in anterior mesenchyme around grafted python cells. Red arrow in **g** (ventral view) indicates anteriorly extended domain of *Fgf8* in chick ectoderm overlying the python mesenchyme cells (compare anterior limit of *Fgf8* expression in limb containing graft (left) with contralateral limb (right)). **h–k**, Dorsoventral polarity in python limb buds. Antibody staining of EN and LMX1 in python limb buds at 1 day of incubation and chick limb buds at stage 20. **h, i**, Expression of EN is detected in ventral ectoderm of python (**h**) and chick (**i**) limb buds. EN expression is also seen in python somites (**h**). Expression of LMX1 is detected in dorsal mesenchyme of python (**j**) and chick (**k**) limb buds. D, dorsal; V, ventral.

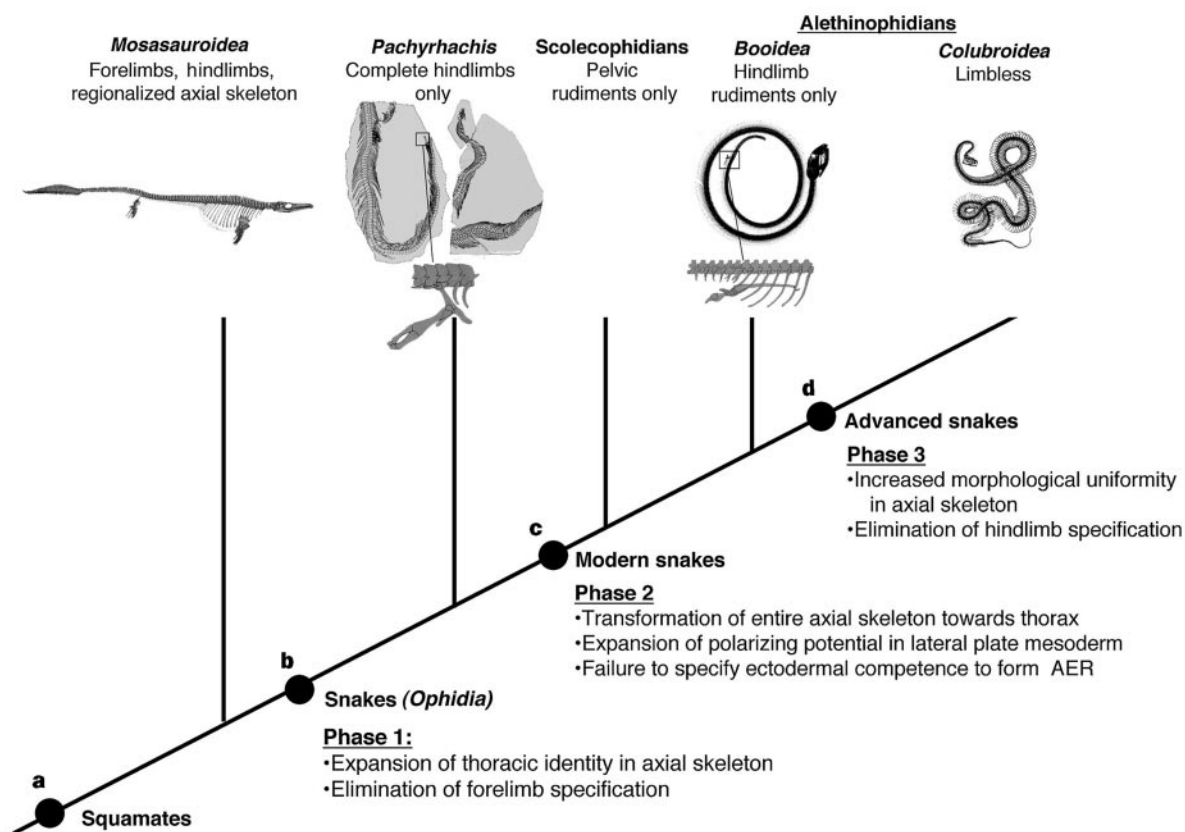


Figure 5 Developmental model for the evolution of snakes. The tree shows the evolutionary relationships among the following: *Colubroidea* (advanced snakes) which lack both forelimbs and hindlimbs and have a large number of nearly-identical vertebrae; *Booidea* (including pythons and boas) which lack forelimbs, but have rudimentary hindlimbs and a large number of morphologically uniform vertebrae with few or equivocal regional differences; Scolecophidians, which have pelvic rudiments and a large number of morphologically uniform vertebrae; the primitive snake *Pachyrhachis problematicus*, which lacks forelimbs, but has complete (or nearly complete) hindlimbs and a large number of similar vertebrae which nonetheless have identifiable regional differences; and mosasaurs, which have a short, morphologically regionalized axial skeleton and complete, normally polarized forelimbs and hindlimbs. According to this model, progressive expansion of Hox gene expression domains can account for loss of forelimbs, hindlimbs and regional identity in the axial skeleton. Additionally, the increase in vertebral number probably required continuous production of

mesoderm for axial elongation, and this could have been achieved by sustained growth of the tail bud and movement of mesoderm through the primitive streak³⁰. Node **a** indicates the origin of squamates. **b**, Hox expansion initiated before the divergence of the *Pachyrhachis* lineage could have led to the reduction of regional differentiation in the axial skeleton and elimination of forelimb specification, with hindlimb development remaining unaffected. **c**, Continued expansion of Hox domains after the divergence of the *Pachyrhachis* lineage could have led to transformation of the entire axial skeleton (anterior to the tail) towards thoracic identity and to reduction of hindlimb development by eliminating ectodermal competence to form an apical ridge and expanding polarizing potential (competence to express *Shh*). This condition is retained in scolecophidians and in modern pythons, which together with boas comprise the *Booidea*. **d**, Further homogenization of Hox gene expression domains is predicted to have led to the origin of advanced snakes/*colubroidea*. (Phylogenetic relationships among these taxa based on ref. 2; figures are modified from refs 32, 33.)

morphological transitions in snake evolution (Fig. 5). Such higher-order genetic changes could have resulted in sudden anatomical transformations, rather than gradual changes, during snake evolution, a hypothesis which can be tested by the fossil record. Our model predicts that embryos of colubroid snakes should show even more homogenization of Hox gene expression domains along the head-to-tail axis than pythons. □

Methods

Whole-mount skeletal preparations. Python embryos to be double-stained for cartilage and bone were fixed in 80% ethanol, then skinned, eviscerated and dehydrated in 96% ethanol. We then incubated them in acetone, rinsed them in 96% ethanol and stained them for 2–6 h in Alcian blue and Alizarin red in 70% ethanol with 5% acetic acid. Embryos were then rinsed in 96% ethanol followed by tap water before clearing in 1% KOH and a graded glycerol series. Embryos stained for cartilage alone were fixed in 5% trichloroacetic acid and stained with Alcian green, then dehydrated in ethanol and cleared in graded glycerol.

Scanning electron microscopy. Embryos were fixed in modified Tyrode's (1% glutaraldehyde) at 4°C. After post-fixation in 1% osmium in 0.1M

phosphate buffer, specimens were dehydrated in graded ethanol, placed in amyl acetate, critical point dried, sputter-coated with gold particles and viewed on a Hitachi S-530 scanning electron microscope.

Whole-mount *in situ* hybridization and immunohistochemistry. Whole-mount *in situ* hybridization was done as described⁵ with digoxigenin-labelled riboprobes for chick *Fgf8* and *patched*. Whole-mount antibody staining was performed as described for HOXC6⁷, DLX³¹, EN²², SHH²⁵, HOXB5²³ and HOXC8²³. These antibodies were raised against *Xenopus* (HOXC6), butterfly (DLX) and mouse (EN, SHH, HOXB5) proteins, and have been shown to recognize the target epitope in a wide range of vertebrates. For immunohistochemistry on frozen sections, embryos were fixed in 4% paraformaldehyde, equilibrated in 30% sucrose, embedded in Tissue-Tek O.C.T. compound and frozen at –80°C. Serial sections were cut at 10 µm and stained using a Vectastain ABC kit according to manufacturer's instructions.

Tissue transplantation and application of FGF2 beads. Python embryos were washed in PBS buffer, then dissected in culture medium. We removed fragments of mesenchyme from specific locations using electrolytically sharpened tungsten needles and fine forceps. Tissue was incubated in 2% trypsin and ectoderm was removed. (All previous steps were carried out on ice.)

We prepared host chick wing buds by lifting the apical ridge away from the anterior mesenchyme to make a loop. Python tissue was transplanted inside the loop and the chick embryos were then reincubated at 38 °C. FGF beads were prepared as described¹⁷. Python eggs were briefly candled to locate embryos and major vessels, then windowed, and membranes and vessels were carefully detached from the inside of the shell to minimize damage. After FGF beads were implanted into a slit at the apex of python limb buds, the python eggs were resealed and incubated at 30 °C in a humidified box.

Received 27 January; accepted 24 March 1999.

- Carroll, R. *Vertebrate Paleontology and Evolution* (Freeman, New York, 1988).
- Lee, M. S. Y. & Caldwell, M. W. Anatomy and relationships of *Pachyrhachis problematicus*, a primitive snake with hindlimbs. *Phil. Trans. R. Soc. Lond. B* **353**, 1521–1552 (1998).
- Gaunt, S. J. Conservation in the Hox code during morphological evolution. *Int. J. Dev. Biol.* **38**, 549–552 (1994).
- Burke, A. C., Nelson, C. E., Morgan, B. A. & Tabin, C. Hox genes and the evolution of vertebrate axial morphology. *Development* **121**, 333–346 (1995).
- Cohn, M. J. *et al.* Hox9 genes and vertebrate limb specification. *Nature* **387**, 97–101 (1997).
- Rancourt, D. E., Tzuzuki, T. & Capocchi, M. R. Genetic interaction between Hoxb-5 and Hoxb-6 is revealed by nonallelic noncomplementation. *Genes Dev.* **9**, 108–122 (1995).
- Oliver, G., Wright, C. V., Hardwicke, J. & De Robertis, E. M. Differential antero-posterior expression of two proteins encoded by a homeobox gene in *Xenopus* and mouse embryos. *EMBO J.* **7**, 3199–3209 (1988).
- Shashikant, C. S. *et al.* Regulation of *Hoxc-8* during mouse embryonic development: identification and characterization of critical elements involved in early neural tube expression. *Development* **121**, 4339–4347 (1995).
- Wall, N. A., Jones, C. M., Hogan, B. L. & Wright, C. V. Expression and modification of Hox 2.1 protein in mouse embryos. *Mech. Dev.* **37**, 111–120 (1992).
- Oliver, G., De Robertis, E. M., Wolpert, L. & Tickle, C. Expression of a homeobox gene in the chick wing bud following application of retinoic acid and grafts of polarizing region tissue. *EMBO J.* **9**, 3093–3099 (1990).
- Nelson, C. E. *et al.* Analysis of Hox gene expression in the chick limb bud. *Development* **122**, 1449–1466 (1996).
- Saunders, J. W. The proximo-distal sequence of origin of the parts of the chick wing and the role of the ectoderm. *J. Exp. Zool.* **108**, 363–402 (1948).
- Fang, H. & Elinson, R. P. Patterns of distal-less gene expression and inductive interactions in the head of the direct developing frog *Eleutherodactylus coqui*. *Dev. Biol.* **179**, 160–172 (1996).
- Ferrari, D. *et al.* The expression pattern of the Distal-less homeobox-containing gene *Dlx-5* in the developing chick limb bud suggests its involvement in apical ectodermal ridge activity, pattern formation, and cartilage differentiation. *Mech. Dev.* **52**, 257–264 (1995).
- Savage, M. P. *et al.* Distribution of FGF-2 suggests it has a role in chick limb bud growth. *Dev. Dyn.* **198**, 159–170 (1993).
- Davidson, D. R., Crawley, A., Hill, R. E. & Tickle, C. Position-dependent expression of two related homeobox genes in developing vertebrate limbs. *Nature* **352**, 429–431 (1991).
- Niswander, L., Jeffrey, S., Martin, G. R. & Tickle, C. A positive feedback loop coordinates growth and patterning in the vertebrate limb. *Nature* **371**, 609–612 (1994).
- Laufer, E., Nelson, C. E., Johnson, R. L., Morgan, B. A. & Tabin, C. Sonic hedgehog and Fgf-4 act through a signaling cascade and feedback loop to integrate growth and patterning of the developing limb bud. *Cell* **79**, 993–1003 (1994).
- Raynaud, A., Kan, P., Bouche, G. & Duprat, A. M. Fibroblast growth factors (FGF-2) and delayed involution of the posterior limbs of the slow-worm embryo (*Anguis fragilis*, L.). *C. R. Acad. Sci.* **318**, 573–578 (1995).
- Altshuler, M., Clarke, J. D. W. & Tickle, C. Dorsal-ventral ectodermal compartments and origin of apical ectodermal ridge in developing chick limb. *Development* **124**, 4547–4546 (1997).
- Zeller, R. & Duboule, D. Dorsal-ventral limb polarity and origin of the ridge: on the fringe of independence? *BioEssays* **19**, 541–546 (1997).
- Davis, C. A., Holmyard, D. P., Millen, K. J. & Joyner, A. L. Examining pattern formation in mouse, chicken and frog embryos with an En-specific antiserum. *Development* **111**, 287–298 (1991).
- Riddle, R. D. *et al.* Induction of LIM homeobox gene *Lmx-1* by WNT7a establishes dorsoventral pattern in the vertebrate limb. *Cell* **83**, 631–640 (1995).
- Chiang, C. *et al.* Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* **383**, 407–413 (1996).
- Marti, E., Takada, R., Bumcrot, D. A., Sasaki, H. & McMahon, A. P. Distribution of Sonic hedgehog peptides in the developing chick and mouse embryo. *Development* **121**, 2537–2547 (1995).
- Marigo, V., Scott, M. P., Johnson, R. L., Goodrich, L. V. & Tabin, C. J. Conservation in hedgehog signaling: induction of a chicken patched homolog by Sonic hedgehog in the developing limb. *Development* **122**, 1225–1233 (1996).
- Yang, Y. *et al.* Relationship between dose, distance and time in *Sonic Hedgehog*-mediated regulation of anteroposterior polarity in the chick limb. *Development* **124**, 4393–4404 (1997).
- Charite, J., de Graaff, W., Shen, S. & Deschamps, J. Ectopic expression of Hoxb-8 causes duplication of the ZPA in the forelimb and homeotic transformation of axial structures. *Cell* **78**, 589–601 (1994).
- Geduspan, J. S. & MacCabe, J. A. Transfer of dorsoventral information from mesoderm to ectoderm at the onset of limb development. *Anat. Rec.* **224**, 79–87 (1989).
- Wilson, V. & Beddington, R. Expression of T protein in the primitive streak is necessary and sufficient for posterior mesoderm movement and somite differentiation. *Dev. Biol.* **192**, 45–58 (1997).
- Panganiban, G., Sebring, A., Nagy, L. & Carroll, S. The development of crustacean limbs and the evolution of arthropods. *Science* **270**, 1363–1366 (1995).
- Gasc, J.-P. Les rapports anatomiques du membre pelvien vestigial chez les squamates serpentiformes. *Bull. Mus. Nat. Hist. 2e Ser.* **38**, 99–110 (1966).
- Eyewitness Encyclopedia of Nature*. (Dorling-Kindersley Multimedia, London, 1995).

Acknowledgements. We thank Drayton Manor Zoo, Edinburgh Zoo, London Zoo, Welsh Mountain Zoo and J. Fletcher for fertile python eggs; M. Caldwell, A. Cohn, S. Evans, M. Ferguson, E. Kochva, M. Lee, C. O. Lovejoy, K. Patel, J. R. D. Stalvey, V. Wilson and L. Wolpert for discussion; M. Turmaine for assistance with scanning electron microscopy; and E. De Robertis, T. Jessell, A. Joyner, G. Martin, A. McMahon, G. Panganiban, C. Tabin, N. Wall, and R. Zeller for reagents. This research was funded by the BBSRC.

Correspondence and requests for materials should be addressed to M.J.C. (e-mail: M.J.Cohn@reading.ac.uk).

The MAPK kinase Pek1 acts as a phosphorylation-dependent molecular switch

Reiko Sugiura*, Takashi Toda†, Susheela Dhut‡, Hisato Shuntoh‡ & Takayoshi Kuno*

* Department of Pharmacology, Kobe University School of Medicine, Kusunoki-cho, Chuo-ku, Kobe 650-0017, Japan

† Cell Regulation Laboratory, Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK

‡ Faculty of Health Science, Kobe University School of Medicine, 7-10-2 Tomogaoka, Suma-ku, Kobe 650-0142, Japan

The mitogen-activated protein kinase (MAPK) pathway is a highly conserved eukaryotic signalling cascade that converts extracellular signals into various outputs, such as cell growth and differentiation^{1–3}. MAPK is phosphorylated and activated by a specific MAPK kinase (MAPKK)⁴: MAPKK is therefore considered to be an activating regulator of MAPK. Pmk1 is a MAPK that regulates cell integrity⁵ and which, with calcineurin phosphatase, antagonizes chloride homeostasis⁶ in fission yeast. We have now identified Pek1, a MAPKK for Pmk1 MAPK. We show here that Pek1, in its unphosphorylated form, acts as a potent negative regulator of Pmk1 MAPK signalling. Mkh1⁷, an upstream MAPKK kinase (MAPKKK), converts Pek1 from being an inhibitor to an activator. Our results indicate that Pek1 has a dual stimulatory and inhibitory function which depends on its phosphorylation

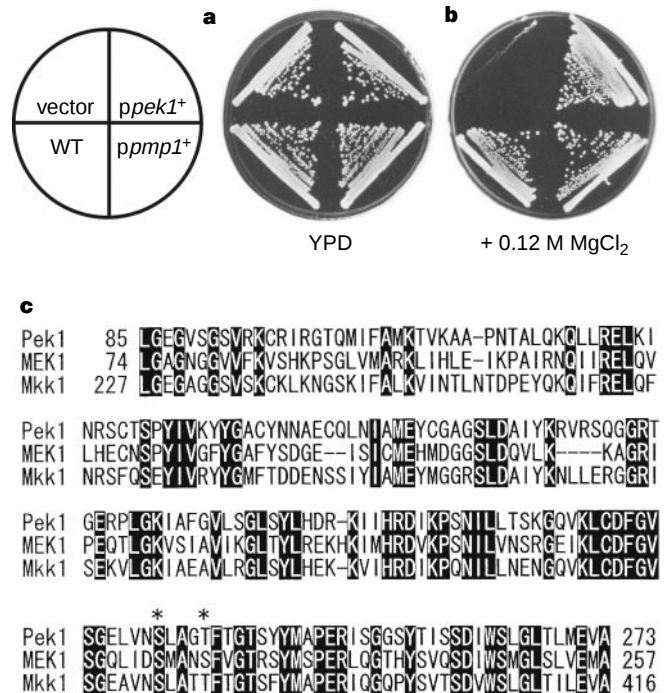


Figure 1 Isolation of Pek1. **a**, **b**, Suppression of the Cl⁻-sensitive growth defect of $\Delta ppb1$. $\Delta ppb1$ cells that had been transformed with multicopy plasmid pDB248 (ref. 25) carrying either *ppb1* (WT), *pek1*⁺ (*ppek1*⁺), *pmp1*⁺ (*ppmp1*⁺), or the vector alone, were streaked onto each plate, containing YPD medium (**a**) or YPD + 0.12 M MgCl₂ (**b**). Plates were incubated for 3 days at 33 °C. **c**, Amino acid sequence alignment of Pek1, human MEK1 (GenBank accession number, L11284), and budding yeast Mkk1 (GenBank accession number, D13001). Filled boxes indicate conserved residues in all three proteins. Asterisks indicate potential phosphorylation sites (S234 and T238).

state. This switch-like mechanism could contribute to the all-or-none physiological response mediated by the MAPK signalling pathway.

Disruption of the calcineurin gene (*ppb1*⁺)⁸ in the fission yeast *Schizosaccharomyces pombe* results in a marked sensitivity to Cl⁻ (ref. 6; Fig. 1a, b). To isolate genes that function in the calcineurin-mediated pathway, we screened an *S. pombe* genome library for genes that, when overexpressed, could suppress the Cl⁻-sensitive growth defect accompanying calcineurin deletion (Δ *ppb1*). One of the known genes is *pmp1*⁺, which encodes a MAPK phosphatase for Pmk1 (ref. 6) (Fig. 1b). In addition to *pmp1*⁺, an unidentified gene rescued the Cl⁻ sensitivity of Δ *ppb1* (Fig. 1b). The nucleotide sequence revealed an open reading frame encoding a polypeptide of 363 amino-acid residues, which shares significant sequence similarity with members of the MAPKK/MEK family^{9,10} (Fig. 1c). We named this gene *pek1*⁺ (for *pombe* MEK 1).

In fission yeast, three distinct MAPK pathways have been identified. These include the mating pheromone-responsive Spk1 MAPK pathway¹¹, the stress-sensing Sty1/Spc1/Phh1 MAPK pathway¹²⁻¹⁴, and the Pmk1/Spm1 MAPK pathway^{5,15} which regulates cell integrity. The *pek1*-deletion mutant (Δ *pek1*) was viable, fertile and formed colonies under high osmolarity (data not shown). However, Δ *pek1* cells exhibited abnormal morphology and cytokinesis similar to Δ *pmk1* cells (Fig. 2a, b, d). Thus, Pek1 is a strong candidate component of the Pmk1 MAPK pathway.

Pek1 is a MAPKK for Pmk1, as indicated by four lines of evidence. First, Δ *pek1* and Δ *pmk1* mutants have essentially identical phenotypes, including abnormal morphology (Fig. 2a, b, d) and a cell-wall

integrity defect (Fig. 2e). Second, these phenotypes are not enhanced in a Δ *pek1* Δ *pmk1* double mutant (Fig. 2c, e). Third, Pek1 regulates the degree of tyrosine phosphorylation in Pmk1; haemagglutinin epitope-tagged Pmk1 (ref. 5) (Pmk1-HA) is tyrosine-phosphorylated in wild-type cells but not in Δ *pek1* cells (Fig. 2f). Moreover, overproduction of a constitutively active mutant, *pek1*^{DD} (see Methods), increased the tyrosine-phosphorylation of Pmk1-HA (Fig. 2f). Fourth, Pek1 binds Pmk1 *in vivo*. A fusion protein between glutathione S-transferase (GST) and Pek1 or unfused GST was co-expressed with Pmk1-HA in *S. pombe* (Fig. 2g), purified with glutathione beads, and analysed by immunoblotting. Pmk1-HA associated with GST-Pek1 (Fig. 2g, lane 4), but not with unfused GST (Fig. 2g, lane 3). Sty1-HA¹² did not associate with GST-Pek1 (data not shown).

Inhibition of Pmk1 MAPK signalling suppresses the effect of calcineurin deletion⁶; disruption of the *pmk1*⁺ gene or of the *mkh1*⁺ gene (encoding MAPKKK), or overexpression of the *pmp1*⁺ gene (encoding MAPK phosphatase) all suppressed the Cl⁻-sensitive growth defect of Δ *ppb1*. Disruption of the *pek1*⁺ gene also suppressed the growth defect (Fig. 3a, c-e).

Thus, the results appear to be conflicting: overexpression of *pek1*⁺ (*pek1*⁺) generated the same phenotype as did its deletion (Δ *pek1*), suggesting that Pek1 can act as an inhibitor as well as an activator. Overexpression of *pek1*⁺ may increase the amount of unphosphorylated Pek1 as well as of phosphorylated Pek1. If the inhibition by unphosphorylated Pek1 overrides the stimulation by phosphorylated Pek1, then overexpression of *pek1*⁺ would inhibit Pmk1 signalling.

To test this idea, we examined the effect of markedly increased wild-type *pek1*⁺ expression by using a pART1 vector¹⁶ containing the strong *adh1* promoter. The introduction of pART1-*pek1*⁺ (*padh-pek1*⁺) suppressed the calcineurin deletion effect more strongly than did *pek1*⁺ overexpression with its natural promoter (*pek1*⁺) (Fig. 3c, d). Thus, the inhibitory effect of Pek1 overproduction on Pmk1 signalling was dose-dependent.

To determine whether the phosphorylation state of Pek1 affected its inhibitory effect, we used Pek1^A (see Methods), which cannot be fully phosphorylated by an upstream MAPKKK. Overproduction of Pek1^A suppressed the Cl⁻ sensitivity arising from calcineurin deletion more effectively than did overproduction of wild-type Pek1

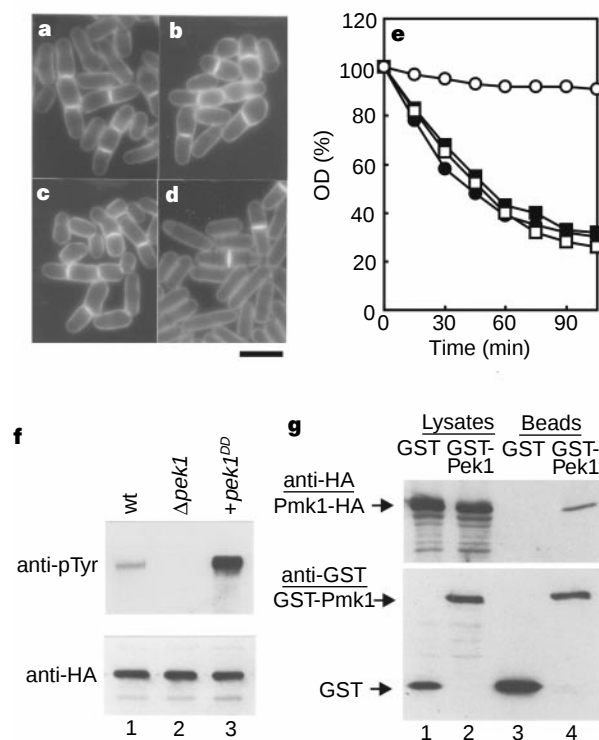


Figure 2 Pek1 is a MAPKK for Pmk1. **a-d**, Cell morphology of Δ *pek1* cells. Δ *pek1* cells (**a**), Δ *pmk1* cells (**b**), or Δ *pek1* Δ *pmk1* cells (**c**) containing the vector; and Δ *pek1* cells carrying *pek1*⁺ (**d**) were cultured in EMM at 33°C. Cells were stained with Calcofluor. Scale bar, 10 μ m. **e**, Cell-wall integrity in Δ *pek1* mutants. Glucanase sensitivity of wild-type (open circles), Δ *pek1* (filled circles), Δ *pmk1* (open squares), Δ *pek1* Δ *pmk1* (filled squares) cells²⁶. **f**, Pek1 regulates tyrosine phosphorylation of Pmk1 *in vivo*. Tyrosine phosphorylation of Pmk1-HA was examined in wild-type cells (lane 1), Δ *pek1* cells (lane 2) and cells overexpressing constitutively active Pek1^{DD} (lane 3). **g**, GST-Pek1 associated with Pmk1-HA *in vivo*. Total cell lysates and proteins bound to glutathione beads were analysed by immunoblotting using anti-HA and anti-GST antibodies.

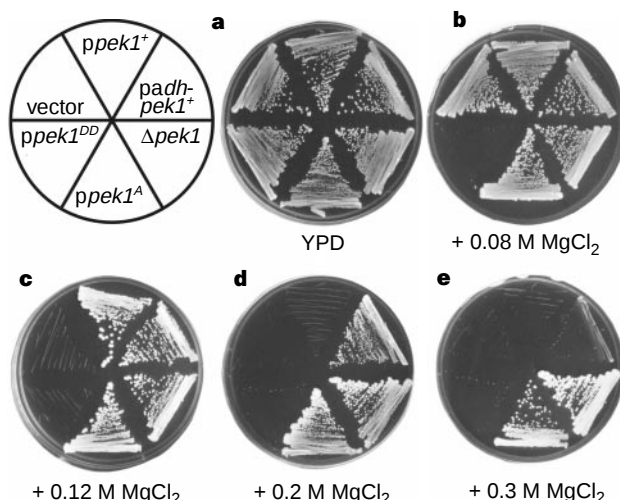


Figure 3 Pek1 has both an inhibitory and stimulatory effect on Pmk1 MAPK signalling. **a-e**, Suppression of Cl⁻-sensitive growth defect of Δ *ppb1* by gene disruption of *pek1*⁺ (Δ *pek1*⁺), or by increased gene dosage of *pek1*⁺ or mutant *pek1* genes at different concentrations of magnesium ion. Δ *pek1* Δ *ppb1* double-mutant cells (Δ *pek1*), and Δ *ppb1* cells transformed with the control plasmid (pDB248), pDB248-*pek1*⁺ (*pek1*⁺), pART1-*pek1*⁺ (*padh-pek1*⁺), pART1-*pek1*^{DD} (*pek1*^{DD}) or pREP1-*pek1*^{DD} (*pek1*^{DD}) were streaked onto the appropriate plate.

(Fig. 3d, e), whereas overexpression of *pek1^{DD}* exacerbated the phenotype. $\Delta ppb1$ strains carrying *pek1^{DD}* failed to grow on media with low concentrations of Cl^- , although $\Delta ppb1$ strains that carry the vector can grow on such media (Fig. 3a, b). These results suggest that Pek1 has a dual inhibitory and stimulatory function which depends on its phosphorylation state.

To rule out the possibility that the dual function of Pek1 could just be an overexpression artefact, we tested mutant alleles of *pek1⁺* expressed at normal cellular concentrations by integrating a single copy of wild-type *pek1⁺*, *pek1^A* or *pek1^{DD}* at the *leu1-32* locus by means of homologous recombination so that it was under the control of the natural promoter. The ratio of expression was 1:1 for pre-existing *pek1⁺* and integrated *pek1⁺*. The integrated single-copy wild-type *pek1⁺* did not alter the calcineurin-deletion effect (Fig. 4a), whereas *pek1^A* suppressed it (Fig. 4a) and *pek1^{DD}* enhanced it (Fig. 4b). Thus, the inhibitory effect of unphosphorylated Pek1 is physiologically significant, and the inhibitory and stimulatory functions of Pek1 are separable.

To see whether Pek1 function depends on Mkh1 activity, we investigated the effect of wild-type and mutant alleles of *pek1⁺* on the defective cell-wall integrity of $\Delta pek1$ cells or $\Delta mkh1$ cells. $\Delta pek1$ cells (Fig. 4c) were fully complemented by wild-type *pek1⁺* or by *pek1^{DD}*, but not by *pek1^A*. In contrast, neither wild-type *pek1⁺* nor

pek1^A had any effect on the results of Mkh1 deletion (Fig. 4d), whereas *pek1^{DD}* suppressed the effect of Mkh1 deletion. We next examined the effect of these *pek1⁺* alleles on the Cl^- sensitivity of $\Delta pek1\Delta ppb1$ cells or $\Delta mkh1\Delta ppb1$ cells (Fig. 4e). In the presence of Mkh1 (Fig. 4e, top), the suppression of the calcineurin-deletion effect by $\Delta pek1$ (vector) was completely reversed by wild-type *pek1⁺*. In the absence of Mkh1 (Fig. 4e, bottom), however, wild-type *pek1⁺* did not have this stimulatory effect, as *pek1⁺* did not affect the suppression by $\Delta mkh1$ (vector); *pek1^{DD}* has a positive effect even in the absence of Mkh1. The stimulatory function of Pek1 evidently requires that it be phosphorylated by Mkh1, indicating that Mkh1 might be able to relieve the inhibition by Pek1 on Pmk1 signalling. We therefore examined the effect of overproducing Mkh1 and Pek1 together on the Cl^- sensitivity of the calcineurin mutant and found that *mkh1⁺* introduced with *pek1⁺* abolished the suppressive effect of *pek1⁺* overexpression (Fig. 5b), indicating that phosphorylation of Pek1 by Mkh1 changes Pek1 from an inhibitor into an activator (Fig. 5c).

Overproduction of Pek1 markedly increases the phosphorylation of Pmk1 (Fig. 6a), so for unphosphorylated Pek1 to exert an inhibitory effect, it must interact with phosphorylated Pmk1 and reduce its activity. We tested the effect of the phosphorylation state of Pek1 on its association with Pmk1 *in vitro* and found that the largest amount of Pmk1-HA was associated with GST-Pek1^A, but only a small fraction of the precipitated Pmk1 was phosphorylated (Fig. 6b); in contrast, co-precipitation of GST-Pek1^{DD} was markedly reduced and precipitated Pmk1 was heavily phosphorylated (Fig. 6b), indicating that phosphorylation of Pek1 reduces its affinity for Pmk1. With wild-type GST-Pek1, a considerable amount of phosphorylated Pmk1 was co-precipitated (Fig. 6b), presumably with Pek1 mainly in its unphosphorylated form.

To test the dual stimulatory and inhibitory function of Pek1, we did reconstitution experiments *in vitro* by purifying mutant forms of Pek1 as GST fusion proteins from fission yeast and examining their effects on Pmk1 kinase activity with myelin basic protein (MBP) as substrate (Fig. 6c). Anti-phospho-Pmk1 antibodies were used to detect Pmk1 phosphorylation (Fig. 6c). Purified GST-Pek1^{DD}

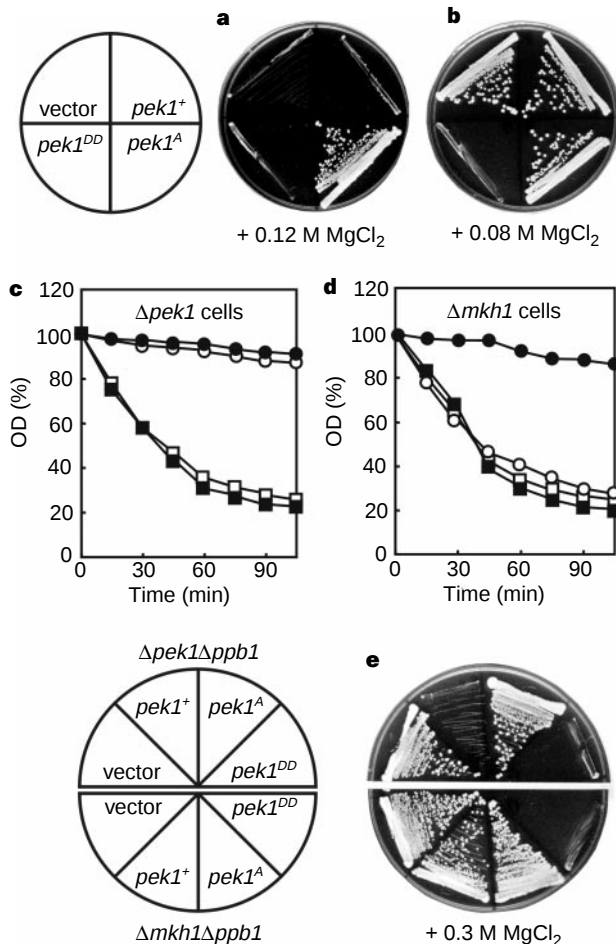


Figure 4 The dual function of Pek1 depends on its phosphorylation state. **a, b**, $\Delta ppb1$ cells integrated with plasmid pJK148 (ref. 22) carrying either wild-type *pek1⁺*, *pek1^A*, *pek1^{DD}*, or vector alone were streaked onto the appropriate plate. **c, d**, The cell wall was digested from $\Delta pek1$ cells (**c**) or $\Delta mkh1$ cells (**d**) carrying plasmid pJK148 integrated with wild-type *pek1⁺* (open circles), *pek1^A* (filled squares), *pek1^{DD}* (filled circles) or vector alone (open squares). **e**, $\Delta pek1\Delta ppb1$ double mutant (top) or $\Delta mkh1\Delta ppb1$ double mutant (bottom) integrated with the same plasmids as in **a** and **b** were streaked onto the appropriate plate.

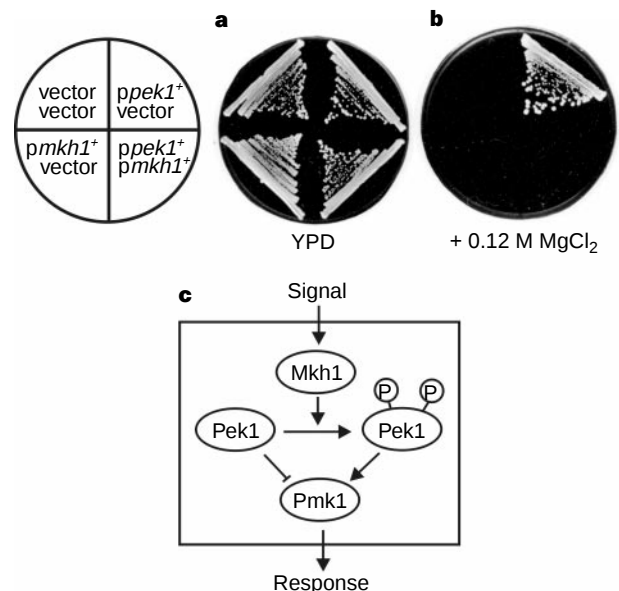


Figure 5 Mkh1 converts Pek1 from an inhibitor to an activator. **a, b**, A calcineurin mutant was transformed with two empty vectors pUR19 (ref. 27) and pDB248 (top left), pUR19-*pek1⁺* and pDB248 (top right), pUR19 and pDB248-*mkh1⁺* (bottom left), or pUR19-*pek1⁺* and pDB248-*mkh1⁺* (bottom right). Transformants were streaked onto the appropriate plate. **c**, Model for Pek1 as a molecular switch: unphosphorylated Pek1 is an inhibitor (line with bar) of Pmk1 signalling; phosphorylation by Mkh1 switches Pek1 from an inhibitor to an activator.

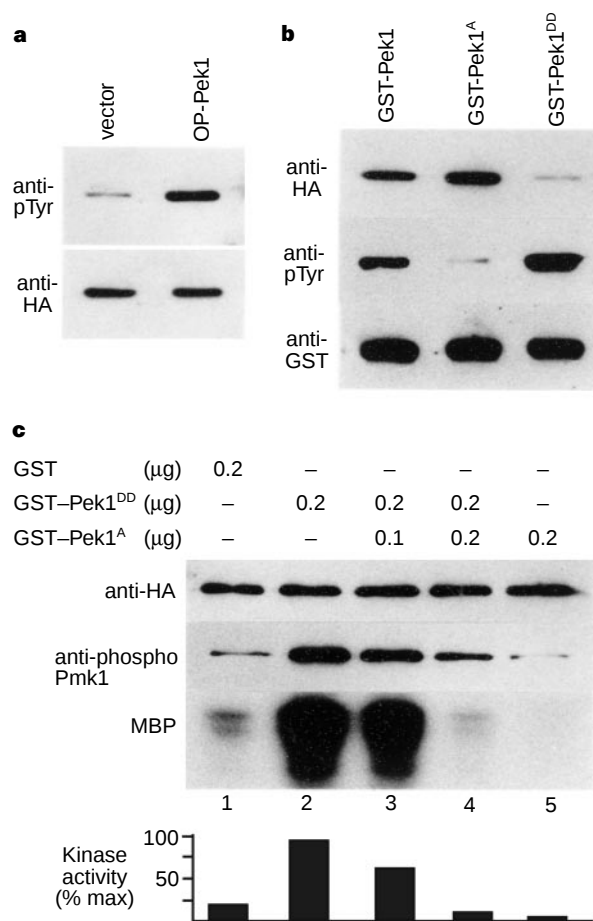


Figure 6 The inhibitory effect of Pek1 on Pmk1 MAPK signalling. **a**, Overproduction of Pek1 increases phosphorylation of Pmk1. Pmk1-HA/6His⁵ was isolated by Ni²⁺-NTA affinity precipitation from cells containing pDB248 (vector) or pDB248-*pek1*⁺ (OP-Pek1), and was immunoblotted using anti-phosphotyrosine (anti-pTyr) or anti-HA antibodies. **b**, Pek1 binds phosphorylated Pmk1 *in vivo*. GST-Pek1, GST-Pek1^A or GST-Pek1^{DD} was co-expressed with Pmk1-HA, purified with glutathione beads, and analysed by immunoblotting using anti-GST, anti-HA or anti-pTyr antibodies. **c**, Unphosphorylated Pek1 inhibits Pmk1 kinase activity *in vitro*. MBP-phosphorylation activity of immunoprecipitated Pmk1-HA was measured in the presence of mutant GST-Pek1 proteins as indicated. Pmk1 kinase activity in lane 2 was set as 100%, and histogram bars correspond to lane numbers. Samples were also immunoblotted by anti-HA or anti-phospho-Pmk1 antibodies.

phosphorylated Pmk1 and markedly increased Pmk1 kinase activity (Fig. 6c, lanes 1, 2). Addition of GST-Pek1^A abolished this activation of Pmk1 kinase by GST-Pek1^{DD} (Fig. 6c, lanes 3–5); in the same sample, a significant amount of Pmk1 remained phosphorylated (Fig. 6c, lane 4). These results indicate that unphosphorylated Pek1 binds phosphorylated Pmk1 and inhibits its kinase activity, thereby attenuating the signalling.

To our knowledge, this is the first demonstration that a MAPK-activating kinase is also a negative regulator of the same MAPK. In budding yeast, the Kss1 MAPK also has a dual function^{17,18}—this protein is phosphorylated and converted from an inhibitor into an activator by an upstream kinase. Thus, such dual-function phosphoproteins may act as molecular switches to turn a continuous input of cellular stimuli into an all-or-none biological response in the MAPK signalling pathway^{19,20}, as well as in other protein-kinase cascades. □

Methods

Gene disruption of *pek1*⁺ and *mkh1*⁺. A 3.2-kb *Bam*HI fragment containing *pek1*⁺ was subcloned into pUC119 to create pUC-pek1. pUC-pek1 was cleaved

at the single *Mlu*I site in *pek1*⁺ and blunt-ended with Klenow polymerase, then ligated to *S. pombe ura4*⁺ (1.8 kb). This construct was used to transform haploid cells²¹. *mkh1*⁺ was disrupted as described⁶.

Expression of Pek1 in yeast. For the expression of Pek1 as a GST fusion protein, a *Bam*HI fragment containing the coding sequence of *pek1*⁺ was amplified by PCR, inserted into the *Bam*HI site of pREP1–GST⁶ to create pREP1–GST–*pek1*⁺. For normal expression of Pek1, we used integrating vector pJK148 (ref. 22) containing *leu1*⁺ as a selectable marker. A *Pst*I–*Not*I fragment containing the endogenous promoter and the coding sequence of *pek1*⁺ amplified by PCR, and a *Not*I–*Sac*I fragment containing triple repeats of HA peptides, were subcloned into the *Pst*I–*Sac*I sites of pJK148 to form pJK148-pek1-HA. pJK148-pek1-HA was linearized at the single *Nru*I site and integrated at the *leu1*-32 locus of each strain. The site-directed mutagenesis of *pek1*^A (Pek1 T238A) or *pek1*^{DD} (Pek1 S234DT238D) was performed using a Sculptor mutagenesis kit (Amersham).

Kinase assays. Pmk1–HA and wild-type or mutant forms of GST–Pek1 were purified from wild-type cells carrying the appropriate plasmid, as described^{23,24}. Kinase reactions consisted of immunoprecipitated Pmk1–HA (0.2 µg), GST-fused Pek1 mutants, and 20 µl incomplete kinase buffer (50 mM Tris–HCl, pH 7.5, 10 mM MgCl₂, 1 mM EGTA, 0.1 mM PMSF, 1% aprotinin, 2 mM benzamide, 1 mM leupeptin) supplemented with 100 µM ATP, 10 µCi [γ -³²P]ATP (Amersham), 0.5 mg ml^{−1} MBP. Reactions were incubated at 30 °C for 30 min, followed by SDS–PAGE, and visualized by autoradiography. Phosphorylated MBP was excised from the gels and quantified. Anti-phospho-Pmk1 antibodies were prepared by immunizing rabbits with a synthetic phosphopeptide (ENPGFMpTEpYVATRWY; amino acids 180 to 194 of Pmk1) conjugated to keyhole limpet haemocyanin by using glutaraldehyde.

Received 9 March; accepted 12 April 1999.

- Herskowitz, I. MAP kinase pathways in yeast: for mating and more. *Cell* **80**, 187–197 (1995).
- Nishida, E. & Gotoh, Y. The MAP kinase cascade is essential for diverse signal transduction pathways. *Trends Biochem. Sci.* **18**, 128–131 (1993).
- Levin, D. E. & Errede, B. The proliferation of MAP kinase signaling pathways in yeast. *Curr. Biol.* **7**, 197–202 (1995).
- Marshall, C. J. MAP kinase kinase kinase, MAP kinase kinase and MAP kinase. *Curr. Opin. Genet. Dev.* **4**, 82–89 (1994).
- Toda, T. *et al.* The fission yeast *pmk1*⁺ gene encodes a novel mitogen-activated protein kinase homolog which regulates cell integrity and functions coordinately with the protein kinase C pathway. *Mol. Cell. Biol.* **16**, 6752–6764 (1996).
- Sugiyama, R. *et al.* *pmp1*⁺, a suppressor of calcineurin deficiency, encodes a novel MAP kinase phosphatase in fission yeast. *EMBO J.* **17**, 140–148 (1998).
- Sengar, A. S., Markley, N. A., Marini, N. J. & Young, D. Mkh1, a MEK kinase required for cell wall integrity and proper response to osmotic and temperature stress in *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* **17**, 3508–3519 (1997).
- Yoshida, T., Toda, T. & Yanagida, M. A calcineurin-like gene *ppb1*⁺ in fission yeast: mutant defects in cytokinesis, cell polarity, mating and spindle pole body positioning. *J. Cell Sci.* **107**, 1725–1735 (1994).
- Crews, C. M., Alessandrini, A. & Erikson, R. L. The primary structure of MEK, a protein kinase that phosphorylates the ERK gene product. *Science* **258**, 478–480 (1992).
- Irie, K. *et al.* MKK1 and MKK2, which encode *Saccharomyces cerevisiae* mitogen-activated protein kinase-kinase homologs, function in the pathway mediated by protein kinase C. *Mol. Cell. Biol.* **13**, 3076–3083 (1993).
- Toda, T., Shimanuki, M. & Yanagida, M. Fission yeast genes that confer resistance to staurosporine encode an AP-1-like transcription factor and a protein kinase related to the mammalian ERK1/MAP2 and budding yeast FUS3 and KSS1 kinases. *Genes Dev.* **5**, 60–73 (1991).
- Millar, J. B., Buck, V. & Wilkinson, M. G. Pyp1 and Pyp2 PTases dephosphorylate an osmosensing MAP kinase controlling cell size at division in fission yeast. *Genes Dev.* **9**, 2117–2130 (1995).
- Shiozaki, K. & Russell, P. Cell-cycle control linked to extracellular environment by MAP kinase pathway in fission yeast. *Nature* **378**, 739–743 (1995).
- Kato, T. *et al.* Stress signal, mediated by a Hog1-like MAP kinase, controls sexual development in fission yeast. *FEBS Lett.* **378**, 207–212 (1996).
- Zaitsevskaia, C. T. & Cooper, J. A. Spm1, a stress-activated MAP kinase that regulates morphogenesis in *S. pombe*. *EMBO J.* **16**, 1318–1331 (1997).
- McLeod, M., Stein, M. & Beach, D. The product of the *mei3*⁺ gene, expressed under control of the mating-type locus, induces meiosis and sporulation in fission yeast. *EMBO J.* **6**, 729–736 (1987).
- Cook, J. G., Bardwell, L. & Thorner, J. Inhibitory and activating functions for MAPK Kss1 in the *S. cerevisiae* filamentous-growth signalling pathway. *Nature* **390**, 85–88 (1997).
- Madhani, H. D., Styles, C. A. & Fink, G. R. MAP kinases with distinct inhibitory functions impart signaling specificity during yeast differentiation. *Cell* **91**, 673–684 (1997).
- Murray, A. W. MAP kinases in meiosis. *Cell* **92**, 157–159 (1998).
- Ferrell, J. E. & Machleder, E. M. The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes. *Science* **280**, 895–898 (1998).
- Rothstein, R. J. One-step gene disruption in yeast. *Methods Enzymol.* **101**, 202–211 (1983).
- Keeney, J. B. & Boeke, J. D. Efficient targeted integration at *leu1*-32 and *ura4*-294 in *Schizosaccharomyces pombe*. *Genetics* **136**, 849–856 (1994).
- Neiman, A. M. & Herskowitz, I. Reconstitution of a yeast protein kinase cascade in vitro: Activation of the yeast MEK homologue STE7 by STE11. *Proc. Natl Acad. Sci. USA* **91**, 3398–3402 (1994).
- Zarrov, P., Mazzoni, C. & Mann, C. The SLT2(MPK1) MAP kinase is activated during periods of polarized cell growth in yeast. *EMBO J.* **15**, 83–91 (1996).
- Beach, D., Piper, M. & Nurse, P. Construction of a *Schizosaccharomyces pombe* gene bank in a yeast bacterial shuttle vector and its use to isolate genes by complementation. *Mol. Gen. Genet.* **187**, 326–329 (1982).
- Levin, D. E. & Bishop, M. A putative protein kinase gene (*kin1*⁺) is important for growth polarity in *Schizosaccharomyces pombe*. *Proc. Natl Acad. Sci. USA* **87**, 8272–8276 (1990).

27. Barbet, N., Muriel, W. J. & Carr, A. M. Versatile shuttle vectors and genomic libraries for use with *Schizosaccharomyces pombe*. *Gene* **114**, 59–66 (1992).
 28. Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795–823 (1991).

Acknowledgements. We thank S. O. Sio, Y. Minami and K. Tanaka for comments on the manuscript, M. Yanagida and H. Mukai for helpful suggestions, and Y. Kawabe for technical assistance. This work was supported in part by research grants from the Ministry of Education, Science and Culture of Japan.

Correspondence and requests for materials should be addressed to T.K. (tkuno@kobe-u.ac.jp). The nucleotide sequence of Pek1 has been deposited in DDBJ, EMBL and GenBank under accession number D82023.

Bcl-2 family proteins regulate the release of apoptogenic cytochrome *c* by the mitochondrial channel VDAC

Shigeomi Shimizu*, Masashi Narita & Yoshihide Tsujimoto*

Osaka University Medical School, Biomedical Research Center, Department of Medical Genetics and *CREST of Japan Science and Technology Corporation (JST), 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan

During transduction of an apoptotic (death) signal into the cell, there is an alteration in the permeability of the membranes of the cell's mitochondria, which causes the translocation of the apoptogenic protein cytochrome *c* into the cytoplasm, which in turn activates death-driving proteolytic proteins known as caspases^{1,2}. The Bcl-2 family of proteins, whose members may be anti-apoptotic or pro-apoptotic, regulates cell death by controlling this mitochondrial membrane permeability during apoptosis^{3–5}, but how that is achieved is unclear. Here we create liposomes that carry the mitochondrial porin channel (also called the voltage-dependent anion channel, or VDAC) to show that the recombinant pro-apoptotic proteins Bax and Bak accelerate the opening of VDAC, whereas the anti-apoptotic protein Bcl-x_L closes VDAC by binding to it directly. Bax and Bak allow cytochrome *c* to pass through VDAC out of liposomes, but passage is prevented by Bcl-x_L. In agreement with this, VDAC1-deficient mitochondria from a mutant yeast did not exhibit a Bax/Bak-induced loss in membrane potential and cytochrome *c* release, both of which were inhibited by Bcl-x_L. Our results indicate that the Bcl-2 family of proteins bind to the VDAC in order to regulate the mitochondrial membrane potential and the release of cytochrome *c* during apoptosis.

Bax- and Bak-dependent changes in mitochondrial membrane permeability that lead to loss of membrane potential ($\Delta\psi$) and cytochrome *c* release^{6–8} are mediated by a polypeptide channel called the permeability transition (PT) pore^{6,8,9}, which includes VDAC, the adenine nucleotide translocator (ANT) and cyclophilin D (refs 10, 11). VDAC is an abundant protein in the outer mitochondrial membrane which forms a large voltage-gated pore in planar lipid bilayers¹², and acts as the pathway for the movement of substances in and out of the mitochondrion. Bax and Bak interact with the PT pore^{8,9}. Bax binds to ANT and can sensitize ANT and PT pores in liposomes to atractyloside, an ANT ligand⁹.

During our search for Bcl-x_L-binding proteins in mitochondria using co-immunoprecipitation with an anti-Bcl-x_L antibody, we detected three proteins of relative molecular masses 33K, 18K and 13K (Fig. 1a). Partial amino-acid sequencing of the 33K protein revealed a match with rat VDAC. Co-immunoprecipitation followed by western blot analysis confirmed the interaction between Bcl-x_L and VDAC in isolated mitochondria (Fig. 1b) and also in Bcl-x_L-expressing HepG2 cells (Fig. 1c). No co-immunoprecipitable interaction of Bcl-x_L with Tom 40 (an integral outer-membrane protein) or with F₁-ATPase (an inner-membrane protein) was detected (data not shown), indicating that the interaction between Bcl-x_L and VDAC was specific. As the PT pore is a polypeptide

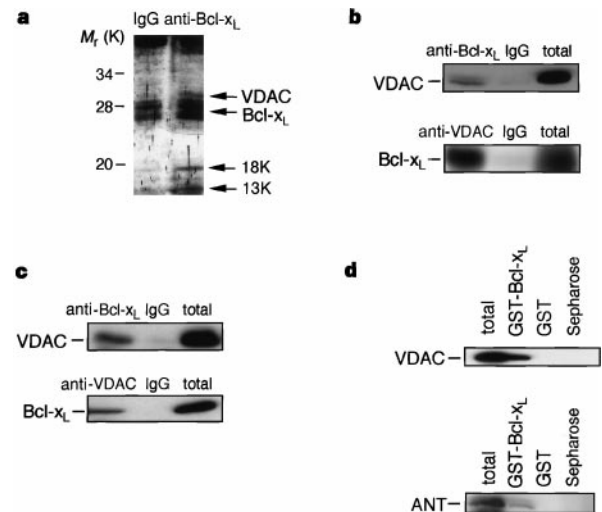


Figure 1 Interaction of Bcl-x_L with VDAC. **a**, Bcl-x_L-binding proteins. Rat liver mitochondria (1 mg ml⁻¹) were incubated with rBcl-x_L (20 mg ml⁻¹) for 5 min and their lysates were immunoprecipitated with anti-Bcl-x_L antibody or rabbit IgG. Immune complexes were analysed by SDS-PAGE and silver staining. **b**, Interaction of rBcl-x_L with VDAC in mitochondria. Mitochondria were incubated with rBcl-x_L, lysed and immunoprecipitated as in **a**, but also with anti-VDAC monoclonal antibody or normal mouse IgG. Anti-VDAC and anti-Bcl-x_L antibodies were used for western blotting. An aliquot of mitochondria (total) was used as the control. **c**, Interaction of rBcl-x_L with VDAC in HepG2 cells. Lysates of human Bcl-x_L-expressing HepG2 cells were immunoprecipitated as for **b**. **d**, Direct interaction between Bcl-x_L and VDAC or ANT. (See Methods for details.) Bound VDAC and ANT were analysed by western blotting; 'total' represents the amount of each protein used.

channel that includes VDAC, co-immunoprecipitation of Bcl-x_L and VDAC from the mitochondria and cells did not necessarily indicate their direct interaction. VDAC purified from rat liver mitochondria efficiently bound to a fusion protein between glutathione-S-transferase and Bcl-x_L (GST-Bcl-x_L), however, indicating that Bcl-x_L interacted directly with VDAC; similar results were obtained when recombinant VDAC was used (data not shown). ANT purified from rat heart bound only weakly to GST-Bcl-x_L (Fig. 1d), consistent with earlier observations⁹.

As Bcl-x_L binds VDAC, we tested the effect of Bcl-x_L on VDAC activity by using VDAC purified from rat liver mitochondria and reconstituted into liposomes. These VDAC liposomes could take up radiolabelled sucrose, whereas plain or heat-denatured VDAC-containing liposomes could not; uptake was greatly reduced in VDAC liposomes preincubated with a polyanion VDAC inhibitor (Fig. 2a), indicating that this sucrose uptake must be mediated by VDAC. Addition of recombinant (r)Bcl-x_L to VDAC liposomes inhibited VDAC-mediated sucrose uptake in a pH-dependent manner (Fig. 2b): Bcl-x_L functioned at acidic pH, consistent with observations that rBcl-x_L acts as an ion channel in synthetic lipid membranes only under acidic conditions^{13–16}. However, when both rBcl-x_L and VDAC were incorporated together into liposomes, Bcl-x_L was able to function at pH 7.3 (Fig. 2c), suggesting that incorporation of Bcl-x_L into the lipid membranes was favoured at lower pH. Although the VDAC has been reported to close below pH 5.0 (ref. 22), our VDAC liposomes showed no difference in behaviour between pH 5.2 and 7.3 (data not shown). VDAC was inhibited by rBcl-x_L in a concentration-dependent manner (Fig. 2d), and two Bcl-x_L mutants (mt-1; F131V, D133A; mt-7; VNW at 135–137 changed to AIL) having partial and no anti-apoptotic activity, respectively¹⁷, gave partial and no inhibition of VDAC-mediated sucrose uptake (Fig. 2d), indicating that those two activities could be linked. As shown in Fig. 2e, mt-1 and mt-7, respectively, bound

VDAC weakly and not at all, indicating that the inhibition of VDAC by Bcl-x_L depends on Bcl-x_L binding to VDAC. VDAC was also inhibited by Bcl-x_L in VDAC liposomes made with recombinant VDAC (Fig. 2f), confirming that VDAC is targeted by Bcl-x_L.

We next tested the effect of Bax on VDAC activity and found that rBax enhanced VDAC-mediated sucrose uptake at pH 5.2 (Fig. 3a), but that uptake of sucrose into plain liposomes incorporating Bax but without VDAC was low (Fig. 3a), indicating that the rBax-enhanced sucrose uptake by VDAC liposomes is mediated not by a putative Bax channel but by the effect of Bax on VDAC. We confirmed this by adding polyanion in the presence of Bax to VDAC liposomes, which significantly inhibited sucrose uptake

(Fig. 3a). As the interaction of VDAC and Bax was not inhibited by polyanion (data not shown), and as polyanion inhibited VDAC activity (Fig. 2a) but not Bax channel activity, as assessed by movement of rhodamine 123 (data not shown), we conclude that Bax enhances VDAC activity. Results were similar for VDAC liposomes made with recombinant VDAC, and when pro-apoptotic Bak (lacking the C-terminal 21 amino-acid residues) was used in place of rBax (Fig. 3b). With liposomes containing both rBax and VDAC, Bax could function at pH 7.3 (data not shown). Consistent with the effect of Bax and Bak on VDAC activity, rBax and rBak could bind purified VDAC (Fig. 3c). Binding between endogenous Bax and VDAC occurred in apoptotic but not intact PC12 cells

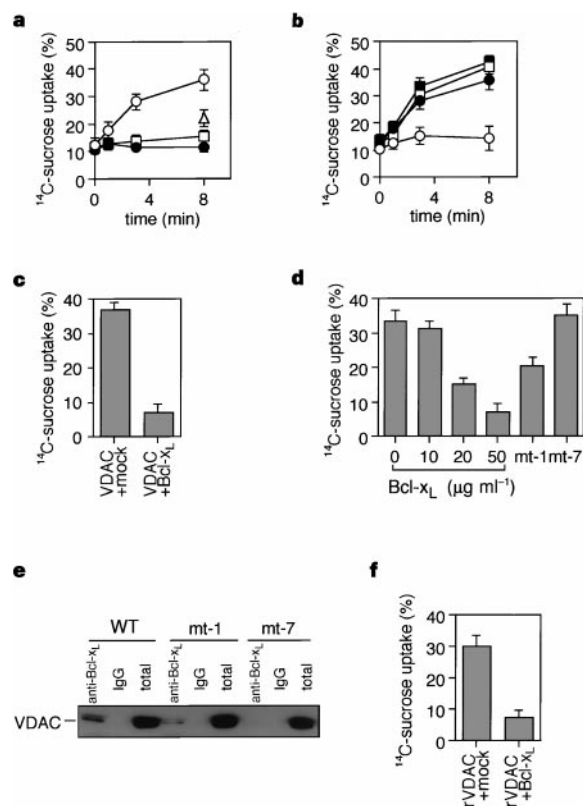


Figure 2 Inhibition of VDAC activity by Bcl-x_L in VDAC liposomes. **a**, Sucrose uptake by VDAC liposomes. Plain liposomes (filled circles), VDAC liposomes (open circles), heat-denatured VDAC liposomes (open squares), and VDAC liposomes with polyanion (100 µg ml⁻¹) (open triangle) were incubated with ¹⁴C-sucrose for the indicated times at pH 5.2 (see Methods). Liposomes were recovered, dissolved in 2% SDS and their ¹⁴C-sucrose content counted. **b**, Inhibition of VDAC activity in VDAC liposomes by Bcl-x_L. VDAC liposomes were incubated with ¹⁴C-sucrose and either rBcl-x_L (20 µg ml⁻¹) (open symbols) or mock protein (filled symbols) at pH 5.2 (circles) and pH 7.3 (squares), and ¹⁴C-sucrose uptake was measured (see Methods). **c**, Inhibition of VDAC activity by Bcl-x_L at neutral pH in liposomes containing VDAC and rBcl-x_L. Purified VDAC and rBcl-x_L were both embedded together into liposomes (see Methods), which were then incubated with ¹⁴C-sucrose at pH 7.3 for 3 min and their ¹⁴C-sucrose uptake measured. **d**, Inhibition of VDAC activity by Bcl-x_L in a dose-dependent manner and effect of Bcl-x_L mutants. VDAC liposomes were incubated with ¹⁴C-sucrose and rBcl-x_L, rBcl-x_L mutant mt-1 or mt-7 protein at 20 µg ml⁻¹ at pH 5.2 for 3 min, and ¹⁴C-sucrose uptake was measured. **e**, Interaction of Bcl-x_L mutants with VDAC. Experiments as in Fig. 1b were done using rBcl-x_L and rBcl-x_L mutants mt-1 and mt-7. Mitochondrial lysates immunoprecipitated with anti-Bcl-x_L antibody or rabbit IgG were western-blotted with anti-VDAC antibody. WT, wild type. **f**, Inhibition of VDAC activity by Bcl-x_L in liposomes incorporating recombinant VDAC. VDAC liposomes were incubated with ¹⁴C-sucrose and rBcl-x_L at pH 5.2 for 3 min and their ¹⁴C-sucrose uptake measured.

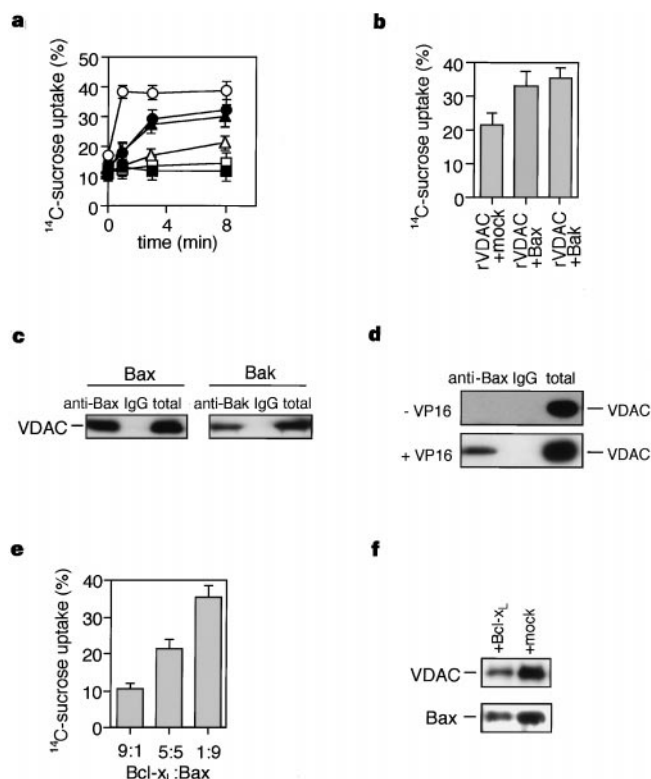


Figure 3 Enhancement of VDAC activity by Bax as well as its inhibition by Bcl-x_L. **a**, ¹⁴C-sucrose uptake into VDAC liposomes is enhanced by Bax. Plain liposomes (squares) and VDAC liposomes (circles and triangles) were incubated with ¹⁴C-sucrose, rBax (20 µg ml⁻¹) (open circles and open squares), rBax (20 µg ml⁻¹) and polyanion (100 µg ml⁻¹) (open triangles), rBaxΔB3 (20 µg ml⁻¹) (filled triangles), and mock protein (filled circles and filled squares) at pH 5.2 for the indicated times before measuring ¹⁴C-sucrose uptake. **b**, VDAC activity is increased by Bax and Bak in liposomes incorporating recombinant VDAC. VDAC liposomes were incubated with ¹⁴C-sucrose and with rBax, rBak or mock protein at pH 5.2 as in **a** and the ¹⁴C-sucrose uptake was counted at 3 min. **c**, Direct interaction of Bax and Bak with VDAC. VDAC liposomes preincubated with rBax or rBak were immunoprecipitated with anti-Bax or anti-Bak antibodies or with rabbit IgG and western-blotted with anti-VDAC antibody. An aliquot of the initial mixture (total) was used as control. **d**, Interaction of endogenous Bax with endogenous VDAC in PC12 cells. Lysates from PC12 cells (a rat pheochromocytoma cell line) preincubated with or without 50 µM VP16 for 12 h were immunoprecipitated with either an anti-mouse Bax antibody that crossreacts with rat Bax or rabbit IgG and the immune complexes western-blotted as in **c**. **e**, Antagonism of sucrose import into VDAC liposomes by Bax and Bcl-x_L. VDAC liposomes were incubated with ¹⁴C-sucrose and with rBax and rBcl-x_L at the indicated ratios (total 20 µg ml⁻¹) at pH 5.2, and ¹⁴C-sucrose uptake measured at 3 min. **f**, Inhibition of interaction with Bax and VDAC by Bcl-x_L. VDAC liposomes incubated with Bcl-x_L or mock protein for 5 min, then with rBax (all at 20 µg ml⁻¹) for 5 min, were immunoprecipitated with anti-Bax (top) or anti-VDAC (bottom) antibodies and western-blotted with anti-Bax or anti-VDAC antibody.

(Fig. 3d). rBax and rBcl-x_L were antagonistic towards VDAC (Fig. 3e). As preincubation of VDAC liposomes with Bcl-x_L significantly inhibited rBax binding to VDAC (Fig. 3f), the effect on VDAC activity of Bax and Bcl-x_L was regulated by their competitive binding to VDAC and their heterodimerization. A recombinant mutant Bax lacking its BH3 region had no effect on VDAC (Fig. 3a), indicating that this region is necessary for enhancing VDAC activity, consistent with our previous observation that the BH3 region is essential for Bak-induced mitochondrial changes⁸. We found that VDAC preparations from rat liver did not contain detectable ANT, as assessed by western blotting, and that sucrose uptake by VDAC liposomes was unaffected by the two ANT ligands atractyloside and bongkreikic acid (data not shown), indicating that we had no contamination by ANT in our VDAC preparations.

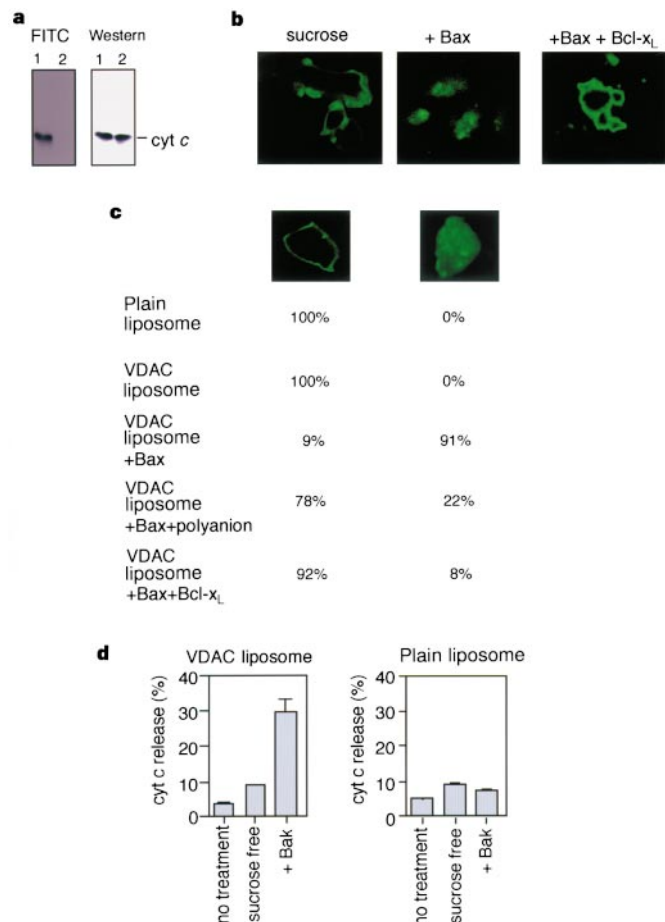


Figure 4 Induction of cytochrome *c* movement through VDAC by Bax and Bak. **a**, FITC labelled cytochrome *c* was run on SDS-PAGE (see Methods) and was detected by fluorescence (left) and western blotting with anti-cytochrome *c* antibody (right): lane 1, FITC-cytochrome *c* (25pg); lane 2, cytochrome *c* (25 pg). **b, c**, Bax-induced import of FITC-cytochrome *c* into VDAC liposomes. VDAC liposomes (**b, c**) and plain liposomes (**c**) were incubated with FITC-cytochrome *c* and sucrose (50 mM), plus rBax (52 µg ml⁻¹), rBax (52 µg ml⁻¹) and polyanion (100 µg ml⁻¹), rBax and rBcl-x_L (both at 52 µg ml⁻¹), or without any of these at pH 5.2 for 5 min (see Methods), and were observed under a confocal fluorescence microscope. Representative photographs (original magnification, ×4,800) of VDAC-liposomes are shown in **b**. In **c**, the frequency (given as a percentage) and representative photographs of liposomes with and without influxed cytochrome *c* are shown (original magnification, ×10,000). **d**, Bak-induced cytochrome *c* release from VDAC liposomes. Plain or VDAC liposomes containing 80 mM sucrose and 0.5 mM cytochrome *c* were incubated in sucrose-free buffer with or without rBak for 30 min at pH 5.2 (see Methods). The release of cytochrome *c* was quantified spectrophotometrically (total incorporated cytochrome *c* is 100%).

As Bax and Bak induce cytochrome *c* release through PT pores of isolated mitochondria^{6–8} and as they enhance VDAC activity, we investigated whether cytochrome *c* could pass through VDAC in the presence of Bax and Bak. To monitor the movement of cytochrome *c* microscopically, we incubated fluorescein isothiocyanate (FITC)-labelled cytochrome *c* (Fig. 4a) with VDAC liposomes. In the absence of rBax, cytochrome *c* accumulated on the surfaces of plain and VDAC liposomes (Fig. 4b, c), but in the presence of rBax (Fig. 4b, c) or rBak (data not shown), cytochrome *c* accumulated inside VDAC liposomes (Fig. 4b, c). Also, polyanion significantly inhibited Bax-induced cytochrome *c* movement (Fig. 4c), indicating that Bax and Bak allowed cytochrome *c* to pass through VDAC. Bax-induced movement of cytochrome *c* was inhibited in the presence of rBcl-x_L (Fig. 4b, c). When cytochrome *c* and sucrose were initially incorporated inside VDAC liposomes, and the liposomes were then incubated with buffer, a significant amount of cytochrome *c* was released only in the presence of rBak (Fig. 4d) and rBax (data not shown); cytochrome *c* was not released by plain liposomes in the presence of rBak (Fig. 4d). Bax/Bak-dependent passage of cytochrome *c* through VDAC occurred also in the absence of sucrose (data not shown). During these experiments, the number of liposomes did not change, as assessed by flow cytometry (data not shown), and a GST-GFP (green fluorescence protein) fusion protein (*M_r* = 50K) was not released when rBax was added to VDAC liposomes containing GST-GFP (data not shown), indicating that the liposomes did not rupture and that Bax and Bak enable cytochrome *c* to pass through VDAC.

Our results indicate that Bcl-2 family proteins may target VDAC directly in order to modulate channel activity. We therefore investigated whether VDAC could be a target for Bcl-2 family proteins that are involved in regulating the apoptosis-associated loss of mitochondrial membrane potential ($\Delta\psi$) and the release of cytochrome *c*. As no mammalian cells lacking VDAC were available and because Bcl-2 family proteins can function in yeast cells^{18,19}, we used a yeast mutant deficient in VDAC1 (Δ VDAC)²⁰ and found that mitochondria isolated from the wild-type yeast lost $\Delta\psi$ and released cytochrome *c* after addition of rBax in a dose-dependent manner (Fig. 5a, d) like mammalian mitochondria, whereas rBax did not cause a loss of $\Delta\psi$ and cytochrome *c* release in mitochondria isolated from Δ VDAC yeast cells (Fig. 5b, d). However, rBax did induce both effects in mitochondria from Δ VDAC yeast transfected with the human (*h*) *vdac1* gene (Fig. 5c, d), indicating that the failure of Δ VDAC yeast mitochondria to respond to rBax was due to the absence of functional VDAC. Bax-mediated $\Delta\psi$ loss and cytochrome *c* release in yeast mitochondria were inhibited by rBcl-x_L (Fig. 5a, c, d). Results were similar when rBak was used (Fig. 5a–c, and data not shown). These results indicate that Bcl-2 family proteins target VDAC to regulate apoptosis-associated mitochondrial loss of $\Delta\psi$ and release of cytochrome *c*.

The fact that Bax and Bak enable cytochrome *c* to pass through VDAC, whereas Bcl-x_L does not, may explain how cytochrome *c* is released during apoptosis. We used the VDAC1-deficient yeast mutant²⁰ to show that VDAC is essential for Bax/Bak-induced mitochondrial $\Delta\psi$ loss and cytochrome *c* release. Although yeast cells carry VDAC1 and VDAC2, VDAC2 makes only a small contribution to the membrane permeability to metabolites.²¹ Extrapolation of these observations suggests that VDAC is essential for cytochrome *c* release during apoptosis of mammalian cells. The estimated size of the VDAC pore is ~2.6 nm, which is not large enough for folded cytochrome *c* with a similar diameter to pass through VDAC²²; indeed, VDAC liposomes did not allow the passage of cytochrome *c* (Fig. 4). Therefore Bax/Bak probably induces a conformational change in VDAC so that cytochrome *c* can pass through the channel, although it is possible that Bax/Bak and VDAC could together form a larger channel, and other unknown mechanisms may contribute to apoptotic cytochrome *c* release *in vivo*.

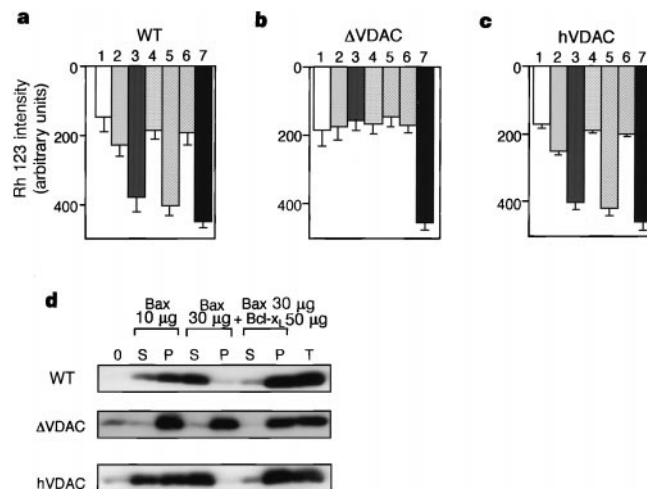


Figure 5 Requirement of VDAC for Bax/Bak-induced loss of membrane potential and cytochrome *c* release in yeast mitochondria. **a–c**, Inability of VDAC-deficient mitochondria to show rBax/rBak-induced loss of membrane potential. Mitochondria (1.0 mg ml^{-1}) isolated from wild-type (WT) yeast (**a**), VDAC1-deficient yeast (ΔVDAC) (**b**), or human VDAC1-expressing ΔVDAC (hVDAC) (**c**) yeast were incubated with: lane 2, $10 \mu\text{g ml}^{-1}$, and lanes 3, 4, $30 \mu\text{g ml}^{-1}$ rBax; lanes 5, 6, $30 \mu\text{g ml}^{-1}$ rBak; lane 1, mock, in the presence (lanes 4, 6) or absence (lanes 3, 5) of $50 \mu\text{g ml}^{-1}$ of Bcl-x_L for 10 min, and $\Delta\psi$ was measured by rhodamine (Rh) 123 intensity; lane 7, complete loss of $\Delta\psi$ was demonstrated by incubation of

mitochondria with 1 mM CCCP for 10 min. **d**, Failure of VDAC1-deficient mitochondria to show Bax-induced cytochrome *c* release. Mitochondria (1 mg ml^{-1}) isolated from wild-type, ΔVDAC or hVDAC1-expressing ΔVDAC yeasts (hVDAC) were incubated with rBax (10 and $30 \mu\text{g ml}^{-1}$) either with or without rBcl-x_L ($50 \mu\text{g ml}^{-1}$) for 10 min, then centrifuged: supernatants (S) and pellets (P) were western-blotted with anti-yeast cytochrome *c* antibody. '0', Supernatant of mitochondria before treatment; 'T', unspun mitochondria. Data are representative of three independent experiments.

Bax and Bcl-2 have been shown to interact with ANT and Bax to sensitize ANT to its atractyloside ligand in ANT liposomes and in partially purified PT-pore liposomes⁹. However, Bax did not affect mitochondrial ADP uptake, which is mediated by ANT (our unpublished observations). Therefore Bax might modulate ANT activity after binding under certain circumstances, but our results strongly indicate that VDAC is a functional target for Bcl-2 family proteins. As Bax/Bak induces loss of $\Delta\psi$ across the inner membrane in mitochondria and as the PT pore is a polyprotein channel, the Bax/Bak-mediated conformational change of VDAC probably affects other channels in the PT pore that control inner-membrane permeability. Alternatively, interaction of Bax/Bak with ANT or ANT-related channels in the inner membrane might be directly involved in $\Delta\psi$ collapse. As PT inhibitors such as cyclosporin A and bongkreic acid can inhibit Bax/Bak-induced cytochrome *c* release in isolated mitochondria^{6,8}, VDAC-mediated cytochrome *c* release could be controlled not only by Bcl-2 family proteins but also by other components of the PT pore such as ANT (a target of bongkreic acid) and cyclophilin D (a target of cyclosporin A). Thus, various components of the PT pore seem to interact with each other. The interaction of Bcl-x_L and Bax with ANT may direct these molecules to bind antagonistically with VDAC inside the PT pore rather than with VDAC in isolation.

Taken together, our results indicate that the Bcl-2 family of proteins target VDAC to regulate apoptosis-associated mitochondrial changes that are central to determining the survival or death of cells.

Methods

Antibodies. Anti-pigeon cytochrome *c* monoclonal antibody (7H8.2C12), anti-yeast cytochrome *c* polyclonal antibody and anti-rat ANT polyclonal antibody were kindly provided by E. Margoliash, G. Shatz and H. H. Schmid, respectively. Anti-human VDAC monoclonal antibody (31HL), which cross-reacts with rat VDAC, was from Calbiochem. Anti-human Bcl-x_L (L19), anti-human Bak (G23) and anti-human Bax (N20) polyclonal antibodies were from Santa Cruz Biotechnology.

Immunoprecipitation and western blot analysis. Rat liver mitochondria were isolated as described²³. Immunoprecipitation was done as described⁸ using cells, mitochondria and liposomes lysed and sonicated in lysis buffer (10 mM

HEPES, pH 7.4, 142.5 mM KCl, 5 mM MgCl₂, 1 mM EGTA, 0.5% N-P40) containing proteinase inhibitors. To detect binding between purified proteins, VDAC ($0.2 \mu\text{g}$) and ANT ($0.2 \mu\text{g}$) were incubated for 3 h either with GST-Bcl-x_L ($0.2 \mu\text{g}$) or GST ($0.2 \mu\text{g}$) in $100 \mu\text{l}$ of the lysis buffer. Then, these proteins were incubated with glutathione (GSH)–Sepharose or Sepharose for 3 h. After brief centrifugation, beads were washed and resuspended in sample buffer for SDS–PAGE.

Protein purification. Human Bcl-x_L, two Bcl-x_L mutants, human Bak (lacking the C-terminal 21 amino-acid residues), human Bax, Bax lacking the BH3 region (Bax ΔBH3), and mock proteins were produced and purified as described⁸. Haemagglutinin (HA)-tagged human VDAC1 was expressed in *Escherichia coli* strain DE3 using the Xpress System (Invitrogen) and purified on an HA column. All proteins except for VDAC were dissolved in 10 mM HEPES–K⁺ (pH 7.4) with 1 mM DTT, and VDAC was in VDAC buffer (10 mM Tris–Cl, pH 7.0, 1 mM EDTA, and 3% Triton-X100).

Rat liver VDAC was purified as described²⁴. Rat heart ANT was purified as described²⁵. The purity of VDAC and ANT was shown to be $>95\%$.

Reconstitution of VDAC in liposomes. Purified and recombinant VDAC were reconstituted in small unilamellar vesicles by a modification of the sonic freeze–thaw procedure²⁶. Briefly, 100 mg phospholipid (soybean, type II-S) was dissolved into 1 ml liposome medium containing 30 mM sodium sulphate and 20 mM Tricine–NaOH (pH 7.3 or pH 5.2) with or without 80 mM sucrose. After sonication, purified or recombinant VDAC protein ($200 \mu\text{g}$) was added with or without rBcl-x_L ($200 \mu\text{g}$) and rBax ($200 \mu\text{g}$), and this suspension was subjected to two freeze–thaw cycles. Heat-denatured VDAC protein was obtained by heating at 95°C for 10 min. Then unilamellar VDAC liposomes were produced by mild sonication and fractionated on Sephadex G-50 columns. Light scattering from each fraction was measured spectrophotometrically at 520 nm , and the fraction containing 700 – 900 arbitrary units per μl was used.

A sucrose-import experiment was performed by measuring ^{14}C -sucrose uptake. Liposomes ($10 \mu\text{l}$) were incubated with $5 \mu\text{l}$ ^{14}C -sucrose (97% ; $200 \mu\text{Ci ml}^{-1}$) at 25°C and filtered by centrifugation using a 30K limiting filter (Millipore) to remove free ^{14}C -sucrose; incorporated ^{14}C -sucrose was measured in a gamma-scintillation counter (Wallac 1414).

Cytochrome *c* translocation experiment. FITC-labelled cytochrome *c* was produced as described²⁷; the FITC/cytochrome *c* molar ratio was 0.47 . FITC-labelled cytochrome *c* was detected with a lumino-image analyser (LAS-1000, Fujifilm). Liposomes ($2 \mu\text{l}$) were incubated with both sucrose (50 mM) and

FITC-labelled cytochrome *c* (50 μ M) and various proteins in a total volume of 29 μ l at pH 5.2; labelled cytochrome *c* was visualized using confocal microphotography (LSM-745, Olympus).

For cytochrome *c* export, liposomes (1 ml) in sucrose (80 mM)-containing buffer were mixed with cytochrome *c* (100 μ M) and freeze-thawed for two cycles. After unilameralization, 20 μ l liposomes were incubated with 1 ml sucrose-free buffer with or without 20 μ g ml⁻¹ rBak for 10 min at 25 °C, then filtered by centrifugation through a 30K filter to separate liposomes from free cytochrome *c*. Free cytochrome *c* was determined from the absorbance at 408 nm.

Yeast mitochondria. AVDAC1-deficient yeast strain (Δ VDAC; M22-2) and its parent strain (M3) were used. A human VDAC1-expressing Δ VDAC yeast strain was produced by transfecting human *vdac1* cDNA using lithium acetate. Mitochondria were isolated essentially as described²⁸, using zymolyase 20T to form spheroplasts which were then homogenized. Isolated mitochondria were suspended in 0.3 M mannitol, 10 mM Tris-maleate (pH 7.4), 0.2 mM EDTA, 0.05% BSA (yMT buffer). Mitochondria (0.5 mg ml⁻¹) were incubated at 25 °C in yMT buffer plus 4.2 mM succinate. $\Delta\psi$ was measured by following the uptake of rhodamine 123 (Rh123) as described²³. For detecting the release of cytochrome *c*, mitochondria were spun, and the supernatants and pellets were western-blotted with anti-yeast cytochrome *c* antibodies.

Received 15 January; accepted 1 April 1999.

- Thornberry, N. A. & Lazebnik, Y. Caspases: Enemies within. *Science* **281**, 1312–1316 (1998).
- Green, D. R. & Reed, J. C. Mitochondria and apoptosis. *Science* **281**, 1309–1312 (1998).
- Merry, D. E. & Korsmeyer, S. J. Bcl-2 gene family in the nervous system. *Annu. Rev. Neurosci.* **20**, 245–267 (1997).
- Adams, J. M. & Cory, S. The Bcl-2 protein family: arbiters of cell survival. *Science* **281**, 1322–1326 (1998).
- Tsujimoto, Y. Role of Bcl-2 family proteins in apoptosis: apoptosomes or mitochondria? *Genes to Cells* **3**, 697–707 (1998).
- Jürgensmeier, J. M. *et al.* Bax directly induces release of cytochrome *c* from isolated mitochondria. *Proc. Natl Acad. Sci. USA* **95**, 4997–5002 (1998).
- Eskes, R. *et al.* Bax-induced cytochrome *c* release from mitochondria is independent of the permeability transition pore but highly dependent on Mg²⁺ ions. *J. Cell Biol.* **143**, 217–224 (1998).
- Narita, M. *et al.* Bax interacts with the permeability transition pore to induce permeability transition and cytochrome *c* release in isolated mitochondria. *Proc. Natl Acad. Sci. USA* **95**, 14681–14686 (1998).
- Marzo, I. *et al.* Bax and adenine nucleotide translocator cooperate in the mitochondrial control of apoptosis. *Science* **281**, 2027–2031 (1998).
- Bernardi, P., Broekemeier, K. M. & Pfeiffer, D. R. Recent progress on regulation of the mitochondrial permeability transition pore: a cyclosporin-sensitive pore in the inner mitochondrial membrane. *J. Bioenerg. Biomembr.* **26**, 509–517 (1994).
- Zoratti, M. & Szabó, I. The mitochondrial permeability transition. *Biochim. Biophys. Acta* **1241**, 139–176 (1995).
- Colombini, M. Voltage gating in the mitochondrial channel, VDAC. *J. Membr. Biol.* **111**, 103–111 (1989).
- Antonsson, B. *et al.* Inhibition of Bax channel-forming activity by Bcl-2. *Science* **277**, 370–372 (1997).
- Minn, A. J. *et al.* Bcl-x_L forms an ion channel in synthetic lipid membranes. *Nature* **385**, 353–357 (1997).
- Schudel, S. L. *et al.* Channel formation by antiapoptotic protein Bcl-2. *Proc. Natl Acad. Sci. USA* **94**, 5113–5118 (1997).
- Schlesinger, P. H. *et al.* Comparison of the ion channel characteristics of proapoptotic Bax and antiapoptotic Bcl-2. *Proc. Natl Acad. Sci. USA* **94**, 11357–11362 (1997).
- Cheng, E. H., Levine, B., Boise, L. H., Thompson, C. B. & Hardwick, J. M. Bax-independent inhibition of apoptosis by Bcl-x_L. *Nature* **379**, 554–556 (1996).
- Greenhalf, W., Stephan, C. & Chaudhuri, B. Role of mitochondria and C-terminal membrane anchor of Bcl-2 in Bax induced growth arrest and mortality in *Saccharomyces cerevisiae*. *FEBS Lett.* **380**, 169–175 (1996).
- Matsuyama, S., Xu, Q., Velours, J. & Reed, J. C. The mitochondrial F₀F₁-ATPase proton pump is required for function of the proapoptotic protein Bax in yeast and mammalian cells. *Mol. Cell* **1**, 327–336 (1998).
- Blachly-Dyson, E., Peng, S., Colombini, M. & Forte, M. Selectivity changes in site-directed mutants of the VDAC ion channel: structural implications. *Science* **247**, 1233–1236 (1990).
- Lee, A. C., Xu, X., Blachly-Dyson, E., Forte, M. & Colombini, M. The role of yeast VDAC genes on the permeability of the mitochondrial outer membrane. *J. Membr. Biol.* **161**, 173–181 (1998).
- Mannella, C. A. Conformational changes in the mitochondrial channel protein, VDAC, and their functional implications. *J. Struct. Biol.* **121**, 207–218 (1998).
- Shimizu, S. *et al.* Bcl-2 prevents apoptotic mitochondrial dysfunction by regulating proton flux. *Proc. Natl Acad. Sci. USA* **95**, 1455–1459 (1998).
- DePinto, V., Prezioso, G. & Palmieri, F. A simple and rapid method for the purification of the mitochondrial porin from mammalian tissues. *Biochem. Biophys. Acta* **905**, 499–502 (1987).
- Rück, A., Dolder, M., Wallimann, T. & Brdiczka, D. Reconstituted adenine nucleotide translocase forms a channel for small molecules comparable to the mitochondrial permeability transition pore. *FEBS Lett.* **426**, 97–101 (1998).
- Báthori, G., Sahin-Tóth, M., Fonyó, A. & Ligeti, E. Transport properties and inhibitor sensitivity of isolated and reconstituted porin differ from those of intact mitochondria. *Biochem. Biophys. Acta* **1145**, 168–176 (1993).
- Kelvin, H. & Fowlkes, B. J. in *Current Protocols in Immunology* 5.3.1–5.3.11 (Wiley Interscience, New York, 1990).
- Daum, G., Bohni, P. C. & Schatz, G. Import of proteins into mitochondria. Cytochrome b2 and cytochrome *c* peroxidase are located in the intermembrane space of yeast mitochondria. *J. Biol. Chem.* **257**, 13028–13033 (1982).

Acknowledgements. We thank E. Margoliash, G. Schatz and H. H. Schmid for antibodies against pigeon cytochrome *c*, yeast cytochrome *c*, and rat ANT, respectively; M. Forte for VDAC1-deficient yeast (M22-2), its wild type (M3), and human *vdac1* cDNA; J. M. Hardwick for Bcl-x_L mutant cDNAs; M. Colombini for the polyanion; T. Chittenden for the expression plasmid for the GST-Bak fusion protein; and

N. Tsujimoto for editorial assistance. This study was supported by grants for Scientific Research on Priority Areas, for Center of Excellence Research and for Scientific Research, from the Ministry of Education, Science, Sports and Culture of Japan.

Correspondence and requests for materials should be addressed to Y.T. (e-mail: tsujimot@gene.med.osaka-u.ac.jp).

Interaction of E1 and hSNF5 proteins stimulates replication of human papillomavirus DNA

Daeyoung Lee*, Hekwang Sohn*, Ganjam V. Kalpana† & Joonho Choe*

* Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Taejeon 305-701, Korea

† Department of Molecular Genetics, Albert Einstein College of Medicine, Bronx, New York 10461, USA

Mammalian viruses often use components of the host's cellular DNA replication machinery to carry out replication of their genomes, which enables these viruses to be used as tools for characterizing factors that are involved in cellular DNA replication. The human papillomavirus (HPV) E1 protein is essential for replication of the virus DNA^{1–3}. Here we identify the cellular factor that participates in viral DNA replication by using a two-hybrid assay⁴ in the yeast *Saccharomyces cerevisiae* and E1 protein as bait. Using this assay, we isolated Ini1/hSNF5 (ref. 5), a component of the SWI/SNF complex which facilitates transcription by altering the structure of chromatin⁶. *In vitro* binding and immunoprecipitation confirmed that E1 interacts directly with

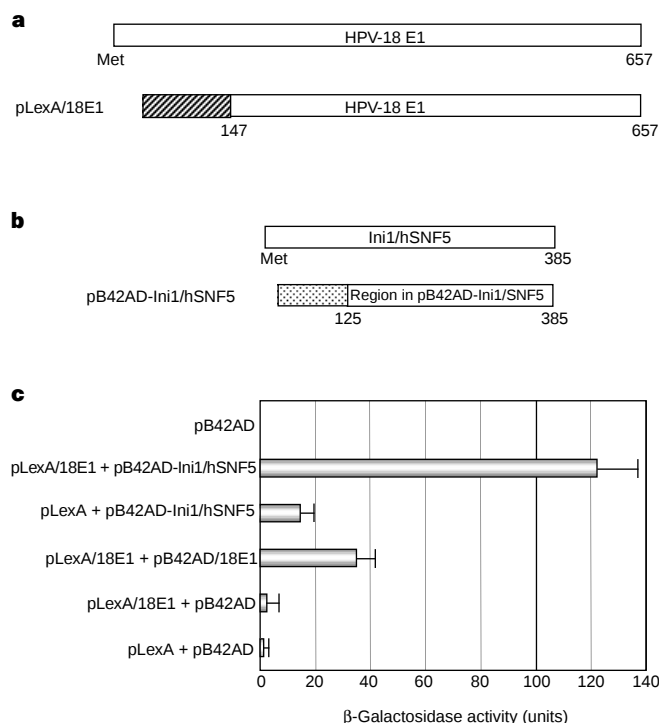


Figure 1 Interaction between HPV-18 E1 and Ini1/hSNF5 proteins. **a**, Structure of the pLexA/18E1 plasmid. Bar represents the LexA DNA binding domain for the LexA operator. **b**, Structure of pB42AD-Ini1/hSNF5. The dotted region represents an activation domain for a yeast promoter. **c**, Quantification of yeast two-hybrid interactions. Values represent the average for five different EGY48 colonies co-expressing the specified fusion proteins. Error bars indicate the standard deviation for each sample.

Ini1/hSNF5. Transient DNA-replication assay revealed that HPV DNA replication is stimulated in a dose-dependent manner by addition of Ini1/hSNF5, and that Ini1/hSNF5 antisense RNA blocks the replication of HPV DNA. Amino-acid substitution at residues that are conserved among E1 proteins prevented the E1-Ini1/hSNF5 interaction and reduced DNA replication of HPV *in vivo*. Our results indicate that Ini1/hSNF5 is required for the efficient replication of papillomavirus DNA and is therefore needed, either alone or in complex with SWI/SNF complex, for mammalian DNA replication as well.

The papillomavirus family consists of small DNA viruses associated with cutaneous and mucosal squamous epithelial lesions in a number of higher vertebrates, including humans. Some human papillomaviruses (HPV) types, such as HPV-16 and HPV-18, are involved in the pathogenesis of certain human anogenital carcinomas, most notably cervical cancer⁷. When the host-cell replication machinery is available, only two viral proteins, E1 and E2, are essential for papillomavirus DNA replication^{1,2,4}. E1 is a nuclear protein of relative molecular mass 72K which has ATPase, DNA helicase and sequence-specific DNA-binding activities; it unwinds the origin of DNA replication (ori)^{8,9} and associates with human DNA polymerase α -primase *in vitro*¹⁰. Like SV40 large T antigen, E1 can initiate origin-specific viral DNA replication *in vitro*^{3,8,9}. E1 also binds to the amino terminus of the papillomavirus E2 protein, which functions as a viral replication factor and transcriptional activator³. Although the E1-E2-ori complex is important for efficient viral DNA replication both *in vivo* and *in vitro*¹¹ the cellular

proteins that are essential for papillomavirus DNA replication have not been well characterized.

To identify cellular proteins that participate in papillomavirus DNA replication, we looked for proteins that interact with HPV-18 E1 by screening a HeLa cell complementary DNA (cDNA) library using the *S. cerevisiae* two-hybrid system⁴. The HeLa cDNA library was cloned into the pB42AD vector, which contains a transcriptional activation domain. To serve as the 'bait', we constructed a plasmid vector (pLexA/18E1) that expressed a fusion protein with the DNA-binding domain of the LexA protein and the carboxy-terminal region (amino acids 147–657) of the HPV-18 E1 protein (Fig. 1a). Transformation of only pLexA/18E1 into yeast strain EGY48 did not activate *lacZ* transcription.

About 5×10^5 colonies were screened in the two-hybrid assay. Twelve positive clones were identified and sequenced. Three of the 12 positive clones contained cDNAs that encoded a truncated version of the Ini1/hSNF5 protein, a component of the SWI/SNF complex⁵ (Fig. 1b). Ini1/hSNF5 is the human homologue of the SNF5 protein of *S. cerevisiae* (ySNF5)¹² and shares, with ySNF5, 33 to 55% overall sequence identity and 41 to 71% similarity in conserved residues. Quantification of the E1-Ini1/hSNF5 interaction in the two-hybrid assay is shown in Fig. 1c.

The multiprotein SWI/SNF complex is essential for the transcriptional activation of several inducible genes¹³ and functions by altering chromatin structure. Genetic and structural data suggest that SWI/SNF disrupts chromatin structure at promoters to allow increased access to the transcription machinery; SWI/SNF proteins mediate interactions between gene-specific transcriptional activators and general transcription factors to help the general transcription machinery to bind to chromatin.

To test whether HPV-18 E1 interacts with Ini1/hSNF5 *in vitro*, we generated an affinity matrix consisting of a glutathione-S-transferase (GST) affinity matrix bound to Ini1/hSNF5. Full-length, radio-labelled HPV-18 E1 was incubated with this matrix, and matrix-bound E1 was analysed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 2a). Binding of E1 to the Ini1/hSNF5 affinity matrix was detectable even at high salt concentrations (1 M NaCl). As expected, E1 did not bind to the control GST affinity matrix even at low salt concentrations (50 mM NaCl). These results confirm that E1 associates with Ini1/hSNF5 and show that this association does not require truncated E1 protein.

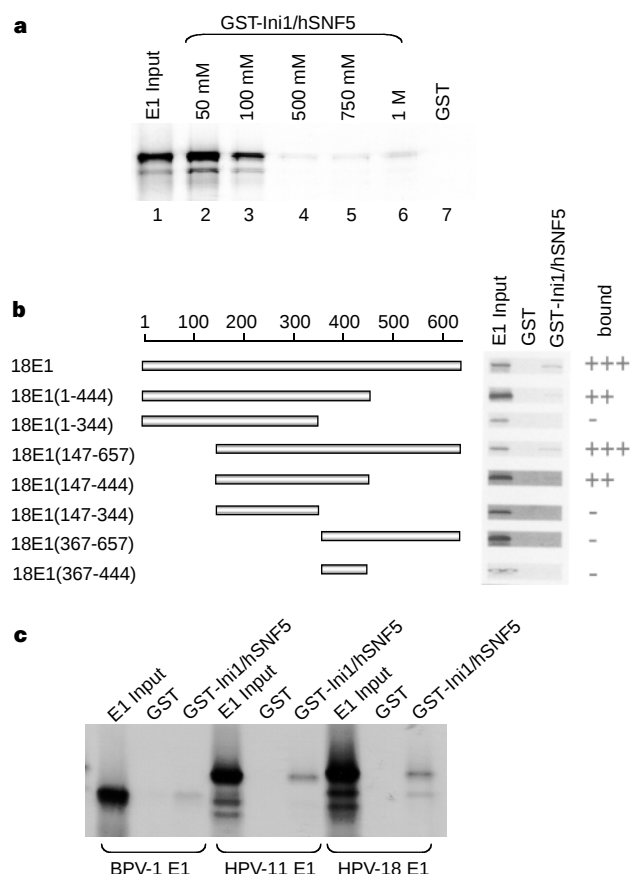


Figure 2 Interaction of HPV and BPV E1 proteins with Ini1/hSNF5. **a**, E1 and Ini1/hSNF5 proteins are bound in various concentrations of salt. E1 input, 25% of input extract. **b**, Various fragments of labelled HPV-18 E1 were incubated with purified GST-Ini1/hSNF5. The relative binding activities are designated with plus and minus signs: +++, 14–8%; ++, 7–4%; +, 3–1%; –, <0.1% binding of the input protein; E1 input, 50% of input extract. **c**, Comparison of the binding activity of Ini1/hSNF5 to E1 from different papillomaviruses.

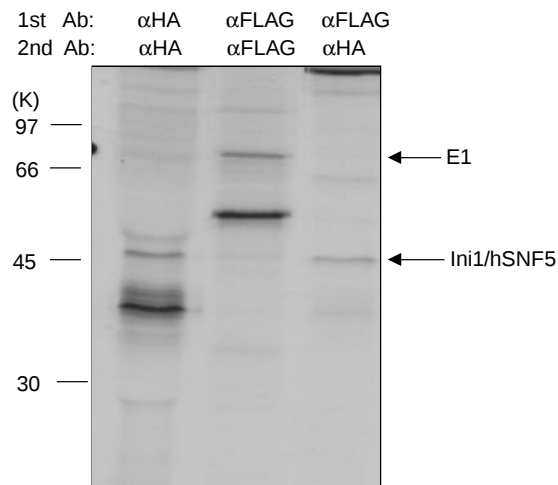


Figure 3 Association of HPV-18 E1 and Ini1/hSNF5 proteins *in vivo*. C33A cells were metabolically labelled with [³⁵S]methionine, and non-denatured extracts were incubated with an anti-Flag monoclonal antibody or anti-HA antibody. The resulting immunoprecipitates were solubilized and re-immunoprecipitated with anti-Flag or anti-HA and precipitates were resolved on 10% SDS-PAGE. Relative molecular mass markers are indicated on the left; E1 and Ini1/hSNF5 proteins are indicated on the right. Ab, antibody; alpha prefix, anti-.

To determine the minimal domain of HPV-18 E1 required for interaction with Ini1/hSNF5, we used the GST-affinity matrix assay to test the binding ability of various E1 deletion mutants (Fig. 2). Full-length HPV-18 E1 (18 E1) and 18E1(147–657) bound tightly to purified GST–Ini1/hSNF5, but not to GST alone; 18E1(1–444) and 18E1(147–444) had moderate Ini1/hSNF5-binding activity. Thus, the region of E1 between residues 147 and 444 constitutes the minimal domain required for binding to full-length Ini1/hSNF5. The defined region does not overlap with the conserved C-terminal domain of E1, which interacts with E2 (ref. 14). These results indicate that E1, E2 and Ini1/hSNF5 form a complex *in vivo* that is important for efficient viral DNA replication. Our preliminary results indicate that E1, E2 and Ini1/hSNF5 form a ternary complex *in vitro* (D.L. and J.C., unpublished results).

We also investigated whether E1 proteins from other papillomaviruses could bind to GST–Ini1/hSNF5 (Fig. 2c) and found that the strength of binding of HPV-11 E1 to GST–Ini1/hSNF5 was comparable to that of HPV-18 E1, and that bovine papillomavirus (BPV-1) E1 bound weakly to GST–Ini1/hSNF5 compared to HPV-18 E1. We conclude that the interaction between papillomavirus E1 and Ini1/hSNF5 is not restricted to HPV-18 E1.

To determine whether E1 and Ini1/hSNF5 associate *in vivo*, we transiently transfected plasmids expressing Flag-tagged HPV-18 E1 and haemagglutinin (HA)-tagged Ini1/hSNF5 into human cervical carcinoma C33A cells. Forty hours after transfection, cells were metabolically labelled with [³⁵S]methionine; non-denatured cell

extracts were then prepared and immunoprecipitated with monoclonal antibody against the HA or Flag epitopes. Immunoprecipitates with anti-HA or anti-Flag were solubilized with SDS, and the denatured proteins were re-immunoprecipitated with anti-Flag or anti-HA. As shown in Fig. 3, the second anti-Flag and anti-HA precipitates contained detectable amounts of E1 and Ini1/hSNF5, respectively. When proteins that immunoprecipitated with E1 and anti-Flag were released from the precipitate and re-immunoprecipitated with anti-HA, HA-tagged Ini1/hSNF5 was recovered, indicating that E1 and Ini1/hSNF5 could be associated in mammalian cells.

We next determined the minimal domain of Ini1/hSNF5 that is required for interaction with HPV-18 E1 by using a β -galactosidase

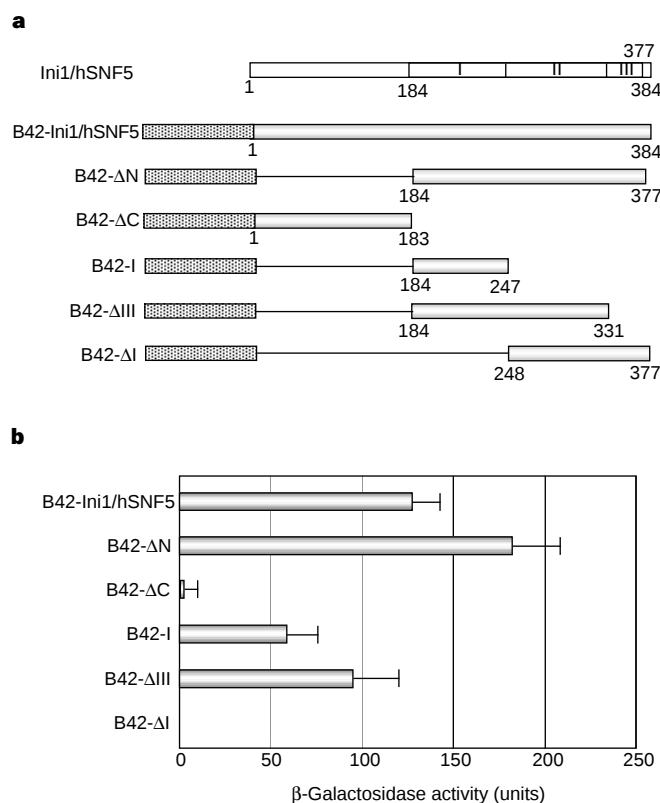


Figure 4 The domain of Ini1/hSNF5 required for binding to full-length HPV-18 E1. **a**, Structure of B42/Ini1/hSNF5 fusion proteins used in yeast two-hybrid assays. We divided the C-terminal part of Ini1/hSNF5 into three regions: regions I and II have a repeated-sequence motif in the conserved Ini1/hSNF5 domain²², and region III contains conserved stretches of charged amino acids¹⁴. Dotted regions represent activation domains for the yeast promoter; grey bars represent segments of Ini1/hSNF5. **b**, β-Galactosidase activity of various Ini1/hSNF5 constructs. Three independent transformants harbouring the plasmids indicated on the left were isolated and grown, and their β-galactosidase activity measured. Values are averages of duplicate assays of three independent transformants, with standard deviations indicated.

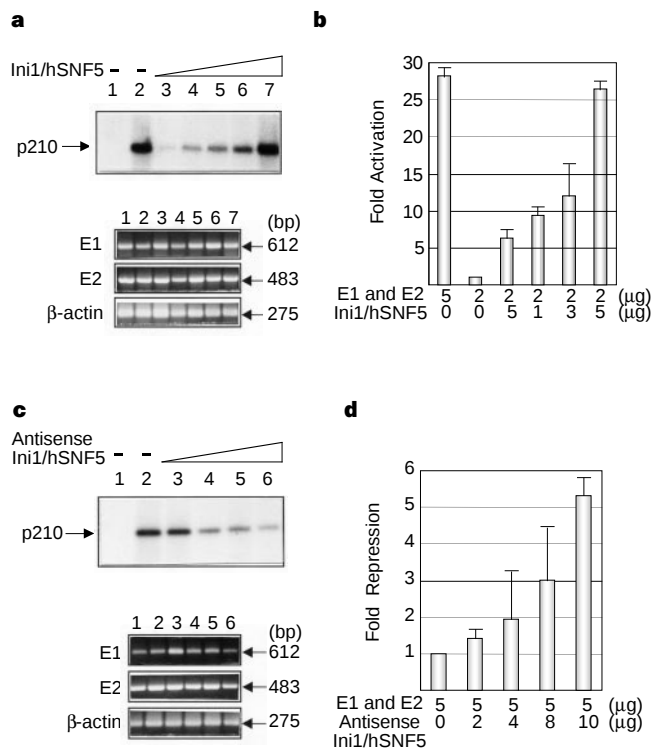


Figure 5 The effects of Ini1/hSNF5 on papillomavirus DNA replication. Transfections are described in Methods. **a**, Activation of *in vivo* DNA replication from the HPV-18 origin by Ini1/hSNF5. Top, E1 and E2 expression vectors (2 μg), p210 (which contains the HPV-18 core origin) (1 μg), and 0, 0.5, 1, 3 or 5 μg pCDNA3-Ini1/hSNF5 (lanes 3 to 7, respectively) was co-transfected into C33A cells. Lane 1, E1 and E2 expression vectors (2 μg); p210 (1 μg) and pBluescriptIIKS(–) were used as a negative control; lane 2, E1 and E2 expression vectors (5 μg) and p210 (1 μg) were used as a positive control. The total amount of transfected DNA was adjusted to 10 μg with sonicated salmon sperm DNA. Bottom, RT-PCR analysis of E1, E2 and β-actin mRNA levels in total RNA preparations from transfected C33A cells. Transcript levels were analysed in parallel with Southern blot analysis. **b**, Quantification of results in **a**. Viral DNA replication is expressed as the degree of relative replication activity (a mixture of 2 μg of E1 and E2 expression vectors and 1 μg of p210 yields one-fold activation). The average of three experiments is shown as relative DNA replication activity, with standard deviation. **c**, Inhibition of HPV-18 origin DNA replication by Ini1/hSNF5 antisense RNA. Top, E1 and E2 expression vectors (5 μg each), p210 (1 μg) and 0, 2, 4, 8 or 10 μg of pCDNA3-antisense-Ini1/hSNF5 (lanes 2 to 8, respectively), or pBluescriptIIKS(–) (lane 1, negative control) were co-transfected into C33A cells. Bottom, RT-PCR analysis of E1, E2 and β-actin mRNA levels in total RNA preparation from transfected C33A cells. Transcript levels were analysed in parallel with Southern blot analysis. **d**, Quantification of results shown in **c**. Viral DNA replication is expressed as the degree of relative replication activity (a mixture of 5 μg of E1 and E2 expression vectors and 1 μg of p210 yields one-fold repression). The average of two experiments is shown as the relative DNA replication activity, with standard deviation.

assay. Yeast strain EGY48 was co-transfected with pLexA/18E1 and pB42AD plasmids containing various fragments of the Ini1/SNF5 gene, and the β -galactosidase activity was measured (Fig. 4a). We found that B42-Ini1/hSNF5 and B42- Δ N interacted strongly and that B42-I and B42- Δ III interacted weakly with HPV-18 E1 compared with control constructs (Fig. 4b). β -Galactosidase activity was very low for B42- Δ C and B42- Δ I. These results indicate that conserved region I (residues 184 to 247)¹⁵ is the minimal domain of Ini1/hSNF5 protein that is required for association with E1. We also observed that the E1-recognition region of Ini1/hSNF5 overlaps with the putative ATPase/helicase hbrm (hSNF2 α)-interaction region within Ini1/hSNF5 (ref. 16). hbrm (hSNF2 α) is the human homologue of yeast SWI/SNF2 and is associated with the human SWI/SNF complex (hSWI/SNF)¹⁷. The SWI/SNF complex facilitates transcription by altering chromatin structure to increase the accessibility of the promoter region for transcription factor binding. The SWI/SNF complex also stimulates nucleosome binding by transcription factors with various DNA-binding domains^{6,18}.

Because changes in chromatin structure are also necessary for DNA replication, we used a transient DNA replication assay to test whether Ini1/hSNF5 affects E1-mediated viral DNA replication (Fig. 5). Plasmid p210, which contains the HPV-18 origin of

replication, was used as a template in the replication assay. DNA replication from the HPV-18 origin was increased in a dose-dependent manner by co-transfection of an Ini1/hSNF5 expression vector (Fig. 5a). It is not clear whether this increase was due to a direct effect of the viral proteins on DNA replication *per se*, or to an Ini1/hSNF5-dependent increase in E1 and E2 proteins *in vivo*. We therefore carried out reverse transcription with the polymerase chain reaction (RT-PCR) of the E1, E2 and β -actin (control) transcripts and found that there was almost the same amount of transcription of E1 and E2 in the presence and absence of the Ini1/hSNF5 expression vector (Fig. 5a, b).

We next tested whether HPV-18 DNA replication was inhibited by Ini1/hSNF5 antisense RNA. If Ini1/hSNF5 participates in viral DNA synthesis, then expression of its antisense RNA should inhibit replication by blocking the synthesis of endogenous Ini1/hSNF5. Increasing amounts of antisense Ini1/hSNF5 RNA expression plasmid were co-transfected into cells, together with the E1 and E2 expression vectors and p210, and viral DNA replication was assayed. Antisense Ini1/hSNF5 RNA inhibited viral DNA replication in a dose-dependent manner (Fig. 5c, d). In addition, RT-PCR analysis of E1 and E2 transcript levels indicated that Ini1/hSNF5 antisense RNA did not affect the expression of the transfected E1 and E2 expression vectors. These results indicate that endogenous Ini1/hSNF5 is required for efficient papillomavirus DNA replication.

To confirm the importance of E1-Ini1/hSNF5 interaction in HPV DNA replication, we introduced point mutations into the Ini1/hSNF5-interacting domain of E1 (Fig. 6a). Amino acids in the Ini1/hSNF5 binding region that are conserved among various papillomavirus E1 proteins were selected for mutagenesis, including Asn 394 and Ala 395, which are also conserved in SV40 large T antigen. Mutations in these amino-acid residues in BPV-1 E1 showed a dominant-negative phenotype in DNA-replication assay¹⁹. All of the E1 mutants tested showed reduced binding to Ini1/hSNF5 in yeast and lower HPV DNA-replication activity compared to wild-type E1 (Fig. 6b, d). Reduced Ini1/hSNF5 binding activity was also confirmed by GST-precipitation assay (Fig. 6c). All E1 mutants retained their ability to bind to E2 in yeast and to stimulate E2-dependent transcription in C33A (data not shown). Our results demonstrate a direct correlation between E1-Ini1/hSNF5 binding and HPV DNA replication. E1 and SV40 large T antigen share amino-acid similarity in eight distinct regions that constitute an important binding domain for ATPase, DNA helicase and DNA polymerase- α (refs 20, 21). An Ini1/hSNF5 binding region may also exist in SV40 large T antigen.

Why does overexpression of Ini1/hSNF5 alone have such marked effect on HPV DNA replication if Ini1/hSNF5 functions as a component of the SWI/SNF complex? Possible explanations for our observations include: Ini1/hSNF5 may affect DNA replication without other members of the complex; the cell may contain inhibitors of the SWI/SNF complex that are titrated out by overexpression of one subunit; and/or Ini1/hSNF5 is the limiting component of the SWI/SNF complex, so its accumulation causes an increase in the activity of the complex.

Proteins that regulate transcription by altering chromatin structure may be important regulators of DNA replication as well. Recruitment of the SWI/SNF complex to the viral origin of replication may trigger unwinding of DNA by means of the E1 DNA helicase activity and thus alter the local arrangement of transcription factors, replication factors and histones. This rearrangement may reprogramme the transcriptional and replicational state of the DNA. E1-mediated *in vivo* DNA replication of BPV-1 is repressed by nucleosomal assembly²², but the transcription factors BPV-1 E2 and NF-1 (ref. 23) alleviate this repression. Cellular proteins such as the SWI/SNF complex may be required for assembly of a replication preinitiation complex in the context of chromatin²². Under such conditions, the interaction of E1 with Ini1/hSNF5 may be a key step in the initiation of viral DNA replication. □

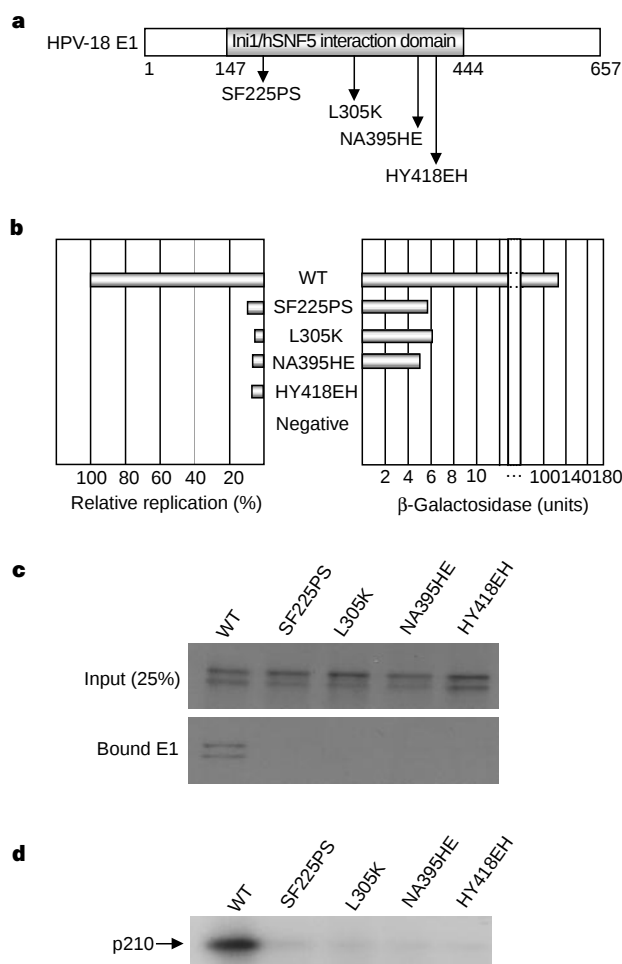


Figure 6 Amino-acid substitutions at conserved residues of E1 proteins prevent E1-Ini1/hSNF5 interaction. **a**, Positions of amino-acid mutations in HPV-18 E1. **b**, Left, *in vivo* DNA replication of wild-type and mutant E1 proteins. Negative, plasmid without the HPV-18 origin. Right, Ini1/hSNF5 binding activity of E1 proteins, determined by β -galactosidase assay in yeast. Negative, pB42AD plasmid, which encodes only a transcriptional activation domain without the Ini1/hSNF5 fragment. DNA replication and β -galactosidase assays were done as for Figs 4 and 5. **c**, GST-pulldown assay of E1 mutant protein binding to Ini1/hSNF5. **d**, E1-driven *in vivo* HPV DNA replication.

Methods

Plasmids. Portions of HPV-18 E1 were generated as *NcoI*–*Sall* restriction fragments by PCR with the appropriate synthetic primers, and were cloned into the pTM1 vector, which contains the EMCV (encephalomyocarditis virus) IRES (internal ribosome entry site) and the T7 promoter. Portions of the yeast expression vector for the In1/hSNF5 deletion mutants fused to the B42 activation domain were generated by subcloning in-frame PCR fragments of In1/hSNF5 into the *EcoRI*–*XhoI* cloning sites of pB42AD. The expression plasmid for HA-In1/hSNF5 was constructed so that the HA epitope was fused to the N terminus of In1/hSNF5. N-terminal, Flag-tagged E1 was cloned into pFLAG-CMV2 (Eastman Kodak) by PCR. The pCDNA3-In1/hSNF5 expression vector was constructed by inserting an *EcoRI*–*XhoI* fragment from the In1/hSNF5 open reading frame (generated by PCR with the appropriate synthetic primers) into pCDNA3 (Invitrogen) linearized with *EcoRI* and *XhoI*. The antisense In1/hSNF5 expression vector pCDNA3-antisense-In1/hSNF5 was constructed by inserting an *EcoRI*–*HindIII* fragment from pUC19-In1/hSNF5 (generated by ligating the *EcoRI*–*XhoI* fragment from pCDNA3-In1/hSNF5 with pUC19 linearized with *EcoRI* and *Sall*) into pCDNA3 linearized with *HindIII* and *EcoRI*.

Yeast two-hybrid assay. The yeast strain EGY48 (*MAT α* *his3* *trp1* *ura3* *LexA_{op(x6)}-LEU3*) harbouring pLexA-18E1 and pSH18-34 was transformed (by the lithium acetate method) with a HeLa cDNA fusion library (20 μ g) cloned into the activation-domain vector pB42AD (Clontech). Transformants were selected by culture on SD/galactose/–Ura/–His/–Trp/–Leu plates for 3 days and were then patched onto fresh SD/glucose/–Ura/–His/–Trp plates. The positive clones were tested for galactose-dependent β -galactosidase activity. β -Galactosidase expression in yeast was assayed as described²⁴.

Metabolic labelling and immunoprecipitation. Cells were labelled for 4 h using 0.5 mCi of [³⁵S]methionine per 10-cm dish in Dulbecco's modified Eagle's medium (DMEM) containing 5% dialysed fetal bovine serum. For detection of E1–In1/hSNF5 immune complexes, about 10⁷ cells were used for each reaction. Radiolabelled cells were lysed for 1 h at 4 °C in 1 ml EBC buffer: 50 mM Tris–HCl at pH 7.5, 120 mM NaCl, 0.5% Nonidet P-40, 50 mM NaF, 200 μ M sodium orthovanadate, 1 mM PMSF. The lysate was centrifuged at 14,000g for 10 min to pellet debris. After preclearing by incubation for 30 min with protein–G–agarose (25 μ l), the supernatant was incubated on a rocker for 1 h with anti-HA or anti-Flag antibody. Precipitates were collected on protein–G–agarose beads, which were washed five times in EBC buffer, resuspended in 50 μ l of SDS lysis buffer (20 mM Tris–HCl at pH 7.5, 50 mM NaCl, 0.5% SDS, 1 mM dithiothreitol), and heated to 95 °C for 2 min. Supernatants were removed and reimmunoprecipitated as described; precipitates were analysed by 10% SDS–PAGE and autoradiography.

RT-PCR. Total RNA was extracted from transfected C33A cells using an SV total-RNA isolation system (Promega). cDNA was synthesized from 500 ng of RNA as described²⁵. PCR was done with 2 μ l cDNA and pairs of oligonucleotide primers designed according to the cDNA sequence, as described²⁵; the product was electrophoresed on a 2% agarose gel and stained with ethidium bromide.

Transient DNA-replication assay. The transient replication assay has been described^{26,27}. Replicated DNA was analysed by agarose-gel electrophoresis and Southern blot hybridization with a random-primed, biotin–dATP-labelled pBluscriptIIKS(–) probe and a chemiluminescent detection method (Phototope-Star detection kit; New England BioLabs).

Mutagenesis. E1 mutants were constructed by using a Quick mutagenesis kit (Stratagene) according to the manufacturer's instructions. Mutants were confirmed by DNA sequencing.

Received 28 January; accepted 31 March 1999.

1. Ustav, M. & Stenlund, A. Transient replication of BPV-1 requires two viral polypeptides encoded by the E1 and E2 open reading frames. *EMBO J.* **10**, 449–457 (1991).
2. Del Vecchio, A. M., Romanczuk, H., Howley, P. M. & Baker, C. C. Transient replication of human papillomavirus DNAs. *J. Virol.* **66**, 5949–5958 (1992).
3. Yang, L., Li, R., Mohr, I. J., Clark, R. & Botchan, M. R. BPV-1 replication *in vitro* is activated by the transcription factor E2. *Nature* **353**, 628–632 (1991).
4. Gyuris, J., Golemis, E., Chertkov, H. & Brent, R. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell* **75**, 791–803 (1993).
5. Kalpana, G. V., Marmon, S., Wang, S., Crabtree, G. R. & Goff, S. P. Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science* **266**, 2002–2006 (1994).
6. Travers, A. A. The reprogramming of transcriptional competence. *Cell* **69**, 573–575 (1992).
7. zur Hausen, H. Papillomavirus infections—a major cause of human cancers. *Biochim. Biophys. Acta* **1288**, F55–F78 (1996).

8. Yang, L. *et al.* The E1 protein of bovine papilloma virus 1 is an ATP-dependent DNA helicase. *Proc. Natl Acad. Sci. USA* **90**, 5086–5090 (1993).
9. Seo, Y. S., Muller, F., Lusk, M. & Hurwitz, J. Bovine papilloma virus (BPV)-encoded E1 protein contains multiple activities required for BPV DNA replication. *Proc. Natl Acad. Sci. USA* **90**, 702–706 (1993).
10. Park, P. *et al.* The cellular DNA polymerase α -primase is required for papillomavirus DNA replication and associates with the viral E1 helicase. *Proc. Natl Acad. Sci. USA* **91**, 8700–8704 (1994).
11. Mohr, I. J. *et al.* Targeting E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. *Science* **250**, 1694–1699 (1990).
12. Laurent, B. C., Treitel, M. A. & Carlson, M. The SNF5 protein of *Saccharomyces cerevisiae* is a glutamine- and proline-rich transcriptional activator that affects expression of a broad spectrum of genes. *Mol. Cell. Biol.* **10**, 5616–5625 (1990).
13. Peterson, C. L. & Herskowitz, I. Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. *Cell* **68**, 573–583 (1992).
14. Yasugi, T., Benson, J. D., Sakai, H., Vidal, M. & Howley, P. M. Mapping and characterization of the interaction domains of human papillomavirus type 16 E1 and E2 proteins. *J. Virol.* **71**, 891–899 (1997).
15. Morozov, A., Yung, E. & Kalpana, G. V. Structure–function analysis of integrase interactor 1/hSNF5L1 reveals differential properties of two repeat motifs present in the highly conserved region. *Proc. Natl Acad. Sci. USA* **95**, 1120–1125 (1998).
16. Muchardt, C., Sardet, C., Bourachot, B., Onufryk, C. & Yaniv, M. A human protein with homology to *Saccharomyces cerevisiae* SNF5 interacts with the potential helicase hhrm. *Nucleic Acids Res.* **23**, 1127–1132 (1995).
17. Kingston, R. E., Bunker, C. A. & Imbalzano, A. N. Repression and activation by multiprotein complexes that alter chromatin structure. *Genes Dev.* **10**, 905–920 (1996).
18. Utley, R. T., Côté, J., Owen-Hughes, T. & Workman, J. L. SWI/SNF stimulates the formation of disparate activator–nucleosome complexes but is partially redundant with cooperative binding. *J. Biol. Chem.* **272**, 12642–12649 (1997).
19. Mansky, K. C., Batiza, A. & Lambert, P. F. Bovine papillomavirus type 1 E1 and simian virus 40 large T antigen share regions of sequence similarity required for multiple functions. *J. Virol.* **71**, 7600–7608 (1997).
20. Clerfant, P. & Seif, I. A common function for polyoma virus large-T and papillomavirus E1 proteins. *Nature* **311**, 276–279 (1984).
21. Masterson, P. J., Stanley, M. A., Lewis, A. P. & Romanos, M. A. A C-terminal helicase domain of the human papillomavirus E1 protein binds E2 and the DNA polymerase α -primase p68 subunit. *J. Virol.* **72**, 7407–7419 (1998).
22. Li, R. & Botchan, M. R. Acidic transcription factors alleviate nucleosome-mediated repression of DNA replication of bovine papillomavirus type 1. *Proc. Natl Acad. Sci. USA* **91**, 7051–7055 (1994).
23. Cheng, L. & Kelly, T. J. Transcriptional activator nuclear factor I stimulates the replication of SV40 minichromosomes *in vivo* and *in vitro*. *Cell* **59**, 541–551 (1989).
24. Guarante, L. Promoters and lacZ fusions designed to study expression of cloned genes in yeast. *Methods Enzymol.* **101**, 181–193 (1983).
25. Gause, W. C. & Adamovics, J. in *PCR Primer: A Laboratory Manual* (eds Dieffenbach, C. W. & Dveksler, G. S.) 293–311 (Cold Spring Harbor Laboratory Press, New York, 1995).
26. Demeret, C., Le Moal, M., Yaniv, M. & Thierry, F. Control of HPV18 DNA replication by cellular and viral transcription factors. *Nucleic Acids Res.* **23**, 4777–4784 (1995).
27. Lee, D., Kim, H., Lee, Y. & Choe, J. Identification of sequence requirement for the origin of DNA replication in human papillomavirus type 18. *Virus Res.* **52**, 97–108 (1997).

Acknowledgements. This work was supported by the Korea Science and Engineering Foundation through the Research Center for Cell Differentiation at Seoul National University and the Academic Research Fund of the Ministry of Education, Republic of Korea. J.C. holds a basic science promotion research grant from KAIST/Samsung Advanced Institute of Technology.

Correspondence and requests for materials should be addressed to J.C. (e-mail: jhchoe@sorak.kaist.ac.kr).

Structure and ligand of a histone acetyltransferase bromodomain

Christophe Dhalluin*, Justin E. Carlson*, Lei Zeng, Cheng He, Aneel K. Aggarwal & Ming-Ming Zhou

Structural Biology Program, Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York, New York 10029-6574, USA

* These authors contributed equally to this work.

Histone acetylation is important in chromatin remodelling and gene activation^{1–4}. Nearly all known histone-acetyltransferase (HAT)-associated transcriptional co-activators contain bromodomains, which are ~110-amino-acid modules found in many chromatin-associated proteins^{5–9}. Despite the wide occurrence of these bromodomains, their three-dimensional structure and binding partners remain unknown. Here we report the solution structure of the bromodomain of the HAT co-activator P/CAF (p300/CBP-associated factor)^{10,11}. The structure reveals an unusual left-handed up-and-down four-helix bundle. In addition, we show by a combination of structural and site-directed mutagenesis studies that bromodomains can interact specifically with acetylated lysine, making them the first known protein modules to do so. The nature of the recognition of acetyl-lysine by the P/CAF

bromodomain is similar to that of acetyl-CoA by histone acetyltransferase. Thus, the bromodomain is functionally linked to the HAT activity of co-activators in the regulation of gene transcription.

The P/CAF bromodomain represents an extensive family of bromodomains (Fig. 1). A large number of long-range nuclear Overhauser enhancement (NOE)-derived distance restraints were identified in the nuclear magnetic resonance (NMR) spectra of the P/CAF bromodomain, yielding a well-defined three-dimensional structure (Fig. 2a, b). The structure consists of a four-helix bundle (helices α_Z , α_A , α_B and α_C) with a left-handed twist, and a long intervening loop between helices α_Z and α_A (termed the ZA loop) (Fig. 2c). The four amphipathic α -helices are packed tightly against one another in an antiparallel manner, with crossing angles for adjacent helices of ~ 16 – 20° . The up-and-down four-helix bundle can adopt two topological folds with opposite handedness (Fig. 2d). The right-handed four-helix bundle fold is more common and is seen in proteins such as haemerythrin and cytochrome b_{562} (refs 12, 13). The left-handed fold of the bromodomain structure is less common, but is also seen in proteins such as cytochrome b_5 and T4 lysozyme^{12,13}. This topological difference arises from the orientation of the loop between the first two helices (Fig. 2d). Right-handed, four-helix-bundle proteins have a relatively short hairpin-like connection between the first two helices, which makes the 'preferred' turn to the right at the top of the first helix^{12–14}. In contrast, proteins with the left-handed fold usually have a long loop after the first helix and often contain additional secondary structural elements at the base of the helix bundle^{12,13}. In the bromodomain structure, this long ZA loop has a defined conformation and is packed against the loop between helices α_B and α_C (termed the BC loop) to form a hydrophobic pocket. These tertiary interactions between the two loops appear to favour the left turn of the ZA loop, resulting in the left-handed, four-helix-bundle fold of the bromodomain. The hydrophobic pocket formed by loops ZA and BC is lined by residues Val 752, Ala 757, Tyr 760, Val 763, Tyr 802 and Tyr 809 (Fig. 2e), and appears to be a site for protein–protein interactions (see below). The pocket is located at one end of the four-helix bundle, opposite the amino and carboxy termini of the protein. The ZA loop varies in length between different bromodomains, but almost always contains residues corresponding to Phe 748, Pro 751, Pro 758, Tyr 760 and Pro 767 (Fig. 1). The conservation of these residues within the ZA loop, as well as residues within the α -helical regions, implies that the large family of bromodomains has a similar left-handed four-helix bundle structure (Fig. 1).

The modular bromodomain structure supports the idea that bromodomains can act as a functional unit for protein–protein interactions⁶. As bromodomains are found in nearly all known nuclear HATs (A-type) that promote transcription-related acetylation of histones on specific lysine residues, but are not present in cytoplasmic HATs (B-type)^{7,8,15,16}, we tested whether bromodomains could interact with acetyl-lysine (AcK). We performed NMR titration of the P/CAF bromodomain with a peptide (SGRGKGG–AcK–GLGK) derived from histone H4, in which Lys 8 is acetylated (Lys 8 is the major acetylation site in H4 for GCN5, a yeast homologue of P/CAF^{1,17}). We found that the bromodomain did bind the AcK peptide, and that this interaction appeared to be specific, on the basis of the ^{15}N -heteronuclear single quantum coherence (HSQC) spectra which showed that only a limited number of residues underwent chemical shift changes as a function of peptide concentration (Fig. 3a). Conversely, NMR titration of the bromodomain with a non-acetylated but otherwise identical H4 peptide showed no noticeable chemical shift changes, demonstrating that the interaction between the bromodomain and the lysine-acetylated H4 peptide depended on acetylation of lysine. The dissociation constant (K_D) for the AcK peptide was estimated to be $346 \pm 54 \mu\text{M}$. This binding is probably reinforced through additional interactions between bromodomain-containing proteins and target proteins. Many chromatin-associated proteins contain two or more bromodomains (Fig. 1)⁶. We also observed binding with another lysine-acetylated peptide (RKSTGG–AcK–APRKQ) derived from the main acetylation site on histone H3 (residues 9–20)¹⁷ (data not shown). Our results show that the P/CAF bromodomain can bind AcK peptides in an acetylation-dependent manner.

The bromodomain residues that exhibited the most significant ^1H and ^{15}N chemical shift changes on peptide binding are located near the hydrophobic pocket between the ZA and BC loops (Fig. 3b). Because the pattern of amide chemical shift changes was similar for the two different AcK-containing peptides, we surmised that the hydrophobic cavity is the primary binding site for AcK. This hypothesis was supported by the results of titration with acetyl-histamine, which mimics the chemical structure of the AcK side chain (Fig. 3c). Both ^{15}N - and ^{13}C -HSQC spectra showed that interaction with acetyl-histamine was also acetylation-dependent, involving the same set of residues that showed chemical shift perturbations with similar concentration dependence (data not shown). The bromodomain did not bind to the amino acids N α -acetyl-lysine or N α -acetyl-histidine alone, possibly because of the

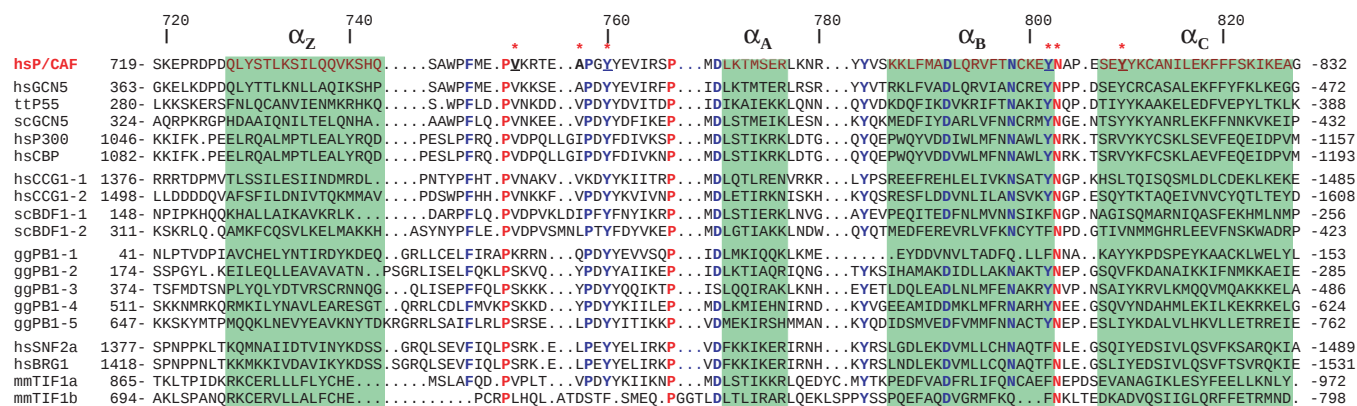


Figure 1 Structure-based sequence alignment of a selected number of bromodomains. The sequences were aligned based on the NMR-derived structure of the P/CAF bromodomain, and the predicted four α -helices are shown in green boxes. Bromodomains are grouped on the basis of the sequence and/or functional similarities as described in ref. 7. Residue numbers of the P/CAF bromodomain are indicated above its sequence. Three absolutely conserved residues, corresponding to Pro 751, Pro 767 and Asn 803 in the P/CAF

bromodomains, are shown in red. Highly conserved residues are coloured in blue. Asterisks show the residues of the P/CAF bromodomain that interact with acetyl-histamine, as determined by intermolecular NOEs. The underlined residues were changed individually by site-directed mutagenesis to alanine. GenBank accession numbers: *hsP/CAF* U57317, *hscGCN5* U57136, *ttP55* U47321, *scGCN5* Q03330, *hsP300* A54277, *hscCBP* S39162, *hscCCG1* P21675, *scBDF1* P3817, *ggPB1* X90849, *hsSNF2a* S45251, *hscBRG1* S39039, *mmTIF1a* S78219, *mmTIF1b* X99644.

charged amino or carboxylic acid group. Particularly, N α -acetyl-histidine has a carboxylate group adjacent to the acetyl moiety (Fig. 3c). These results indicate that the P/CAF bromodomain can interact with acetyl-lysine-containing proteins in a specific manner, and that this interaction is localized to the bromodomain hydrophobic cavity.

To identify the key residues involved in bromodomain-AcK recognition, we determined the NMR structure of the P/CAF

bromodomain in complex with acetyl-histamine. As anticipated, the acetylated moiety binds in the hydrophobic pocket of the bromodomain (Fig. 4). The intermolecular interactions are mainly hydrophobic, with the methyl group of acetyl-histamine making extensive contacts with the side chains of Val 752, Ala 757 and Tyr 760, and the methylene groups of acetyl-histamine displaying specific NOEs to Val 752, Ala 757, Tyr 760, Tyr 802 and Tyr 809. There were no intermolecular NOEs observed for the imidazole ring

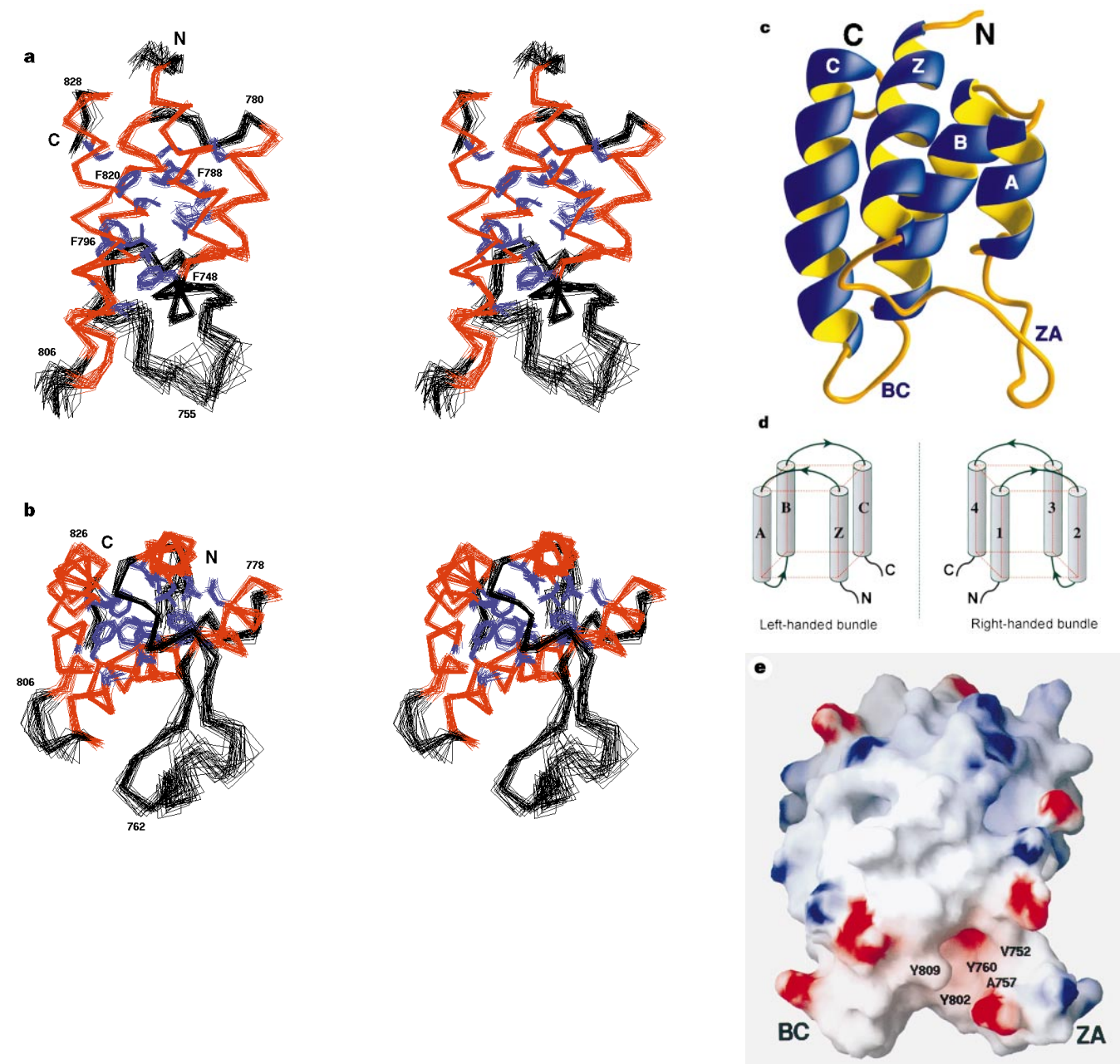


Figure 2 Structure of the P/CAF bromodomain. **a**, Stereoview of the C α trace of 30 superimposed NMR-derived structures of the bromodomain (residues 722-830). The N-terminal four residues (SKEP) that are structurally disordered are omitted for clarity. For the final 30 structures, the root-mean-square deviations (r.m.s.d.s) of the backbone and all heavy atoms are 0.63 ± 0.11 Å and 1.15 ± 0.12 Å for residues 723-830, respectively. The r.m.s.d.s of the backbone and all heavy atoms for the four α -helices (residues 727-743, 770-776, 785-802 and 807-827) are 0.34 ± 0.04 Å and 0.87 ± 0.06 Å, respectively. **b**, Stereoview of the bromodomain structures from the bottom of the protein, which is rotated $\sim 90^\circ$ from the orientation in **a**. **c**, Ribbons²⁹ depiction of the averaged minimized NMR structure

of the P/CAF bromodomain. The orientation of **c** is as shown in **a**. **d**, Schematic representation of the overall topology of the up-and-down four-helix bundle folds with opposite handedness. The left-handed fold is seen in bromodomain, cytochrome b_5 and T4 lysozyme (left), whereas the right-handed four-helix bundles are found in proteins such as haemerythrin and cytochrome b_{562} (right)^{12,13}. **e**, A molecular surface representation of the electrostatic potential (blue, positive; red, negative) of the bromodomain calculated in GRASP³⁰. The hydrophobic and aromatic residues (Tyr 809, Tyr 802, Tyr 760, Ala 757 and Val 752) located between the ZA and BC loops are indicated.

of acetyl-histamine. The spectral analysis shows that the structure of the bromodomain is very similar in both the free and complex forms.

The bromodomain–AcK interaction is similar to that between the histone acetyltransferase Hat1 and acetyl-CoA¹⁸. Although the binding pockets of these two otherwise structurally unrelated proteins are composed of different secondary structural elements,

the nature of acetyl-lysine recognition has striking similarities. In particular, Tyr 809, Tyr 802, Tyr 760 and Val 752 in the bromodomain appear to be related to Phe 220, Phe 261, Val 254 and Ile 217 of Hat1, respectively, in their interactions with the acetyl moiety. This may indicate an evolutionarily convergent mechanism of acetyl-lysine recognition between bromodomains and histone acetyltransferases.

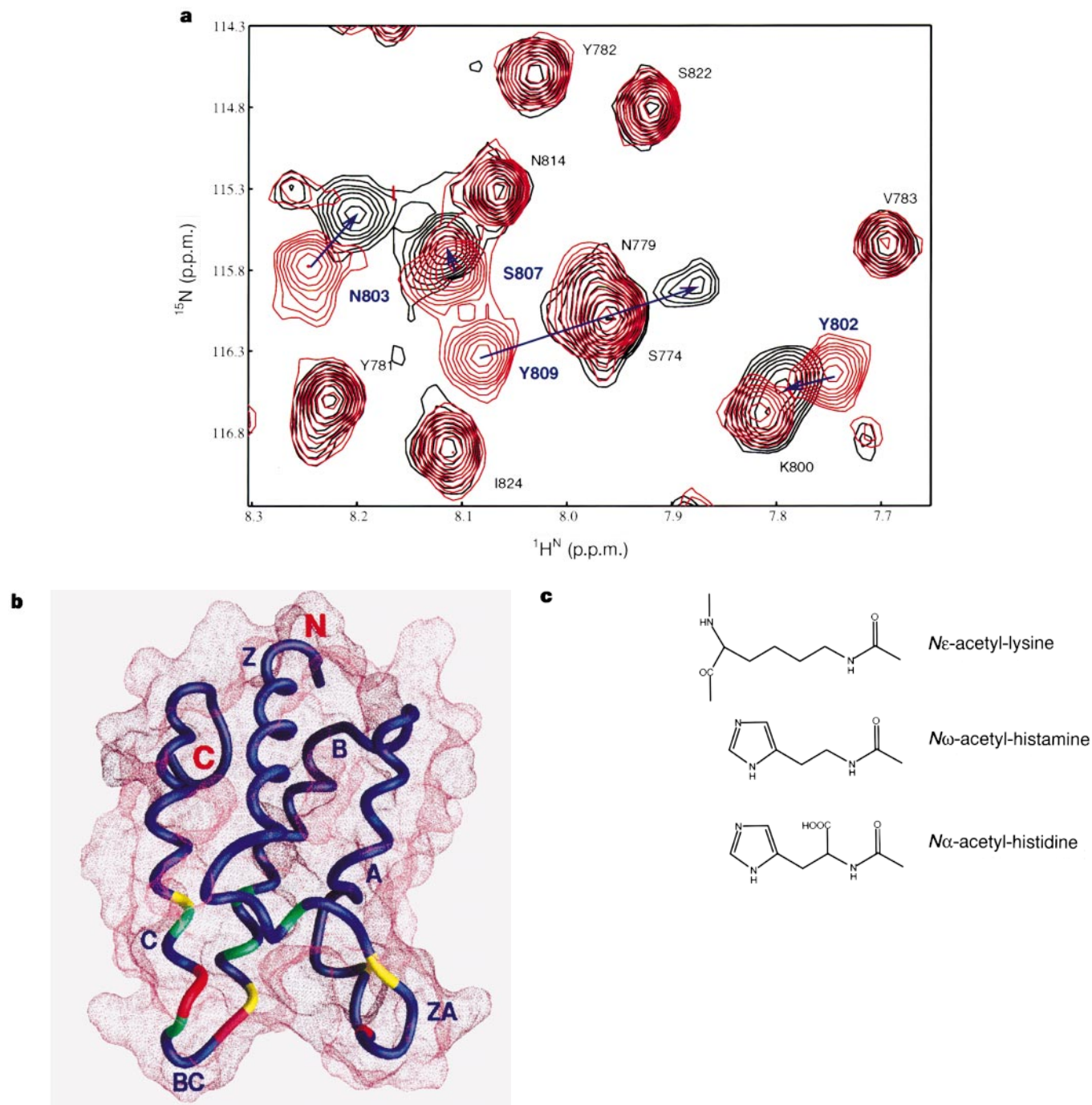


Figure 3 Binding of the P/CAF bromodomain to AcK. **a**, Superimposed region of the 2D ¹⁵N-HSQC spectra of the bromodomain (~0.5 mM) in its free form (red) and complexed to the AcK-containing H4 peptide (molar ratio 1:6) (black). **b**, Ribbon and dotted-surface diagram of the bromodomain depicting the location of the lysine-acetylated H4 peptide-binding site. The colour coding reflects the chemical shift changes ($\Delta\delta$) of the backbone amide ¹H and ¹⁵N resonances upon binding to the AcK peptide as observed in the ¹⁵N-HSQC spectra. The normalized weighted average of the chemical shift changes was calculated by

$\Delta_{av}/\Delta_{max} = [(\Delta\delta_{NH}^2 + \Delta\delta_N^2/25)/2]^{1/2}/\Delta_{max}$, where Δ_{max} is the maximum weighted chemical shift difference observed for Tyr 809 (0.16 ppm). The backbone atoms are colour-coded in red, yellow or green for residues that have Δ_{av}/Δ_{max} of >0.6 (Tyr 809, Glu 808, Asn 803 and Ala 757), 0.2–0.6 (Ala 813, Tyr 802, Tyr 760 and Val 752) or <0.2 (Cys 812, Ser 807, Cys 799, Phe 796 and Phe 748), respectively. The non-perturbed residues are shown in blue. **c**, Chemical structures of acetyl-lysine, acetyl-histamine, and acetyl-histidine.

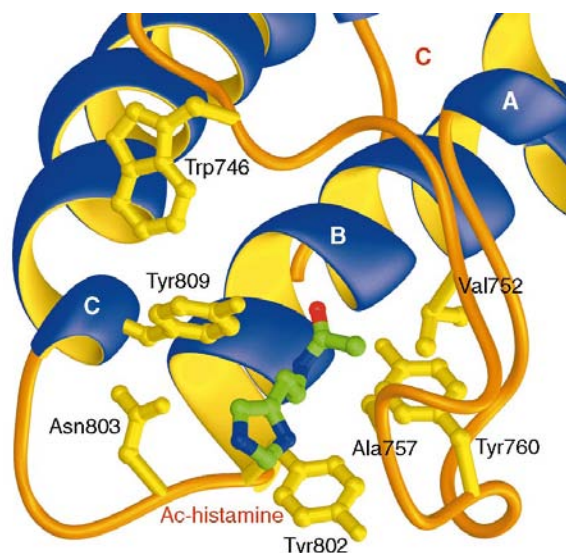


Figure 4 The acetyl-lysine binding pocket. Ribbons²⁹ depiction of a portion of the P/CAF bromodomain complexed with acetyl-histamine. The ligand is colour-coded by atom type.

To determine the relative contributions of residues within the hydrophobic cavity in bromodomain–AcK binding, we used site-directed mutagenesis to alter residues Tyr 809, Tyr 802, Tyr 760 and Val 752 (Table 1). Substitution of alanine for Tyr 809 completely abrogated the bromodomain binding to the lysine-acetylated H4 peptide, whereas the Y802A, Y760A and V752A mutants had significantly reduced ligand-binding affinity. To test whether these mutations disrupted the overall bromodomain fold, we compared the ¹⁵N-HSQC spectra of the mutants to that of the wild-type protein. For the Y809A mutant, the amide chemical shifts were only affected for a few residues near the mutation site. However, mutations of the other residues in the hydrophobic binding pocket caused greater changes in the local protein conformation, particularly in the ZA loop (Table 1). Thus, the results of NMR structural analysis and mutagenesis show that Tyr 809, which is structurally supported by Trp 746 and Asn 803 (Fig. 4), is essential for the bromodomain interaction with the acetyl group of acetyl-lysine, and that Tyr 802, Tyr 760 and Val 752 probably have both structural and functional roles in the recognition. These residues are highly conserved throughout the bromodomain family (Fig. 1), indicating that recognition of acetyl-lysine may be a general feature of bromodomains.

Table 1 Structural and functional analysis of the P/CAF bromodomain mutants

Bromodomain proteins	Structural integrity*	H4 AcK-peptide binding K_D (μ M)†
Wild type	++++	346 ± 54
Y809A	++++	No binding‡
Y802A	+++	>10,000§
Y760A	+++	>10,000
V752A	++	>10,000

* We assessed the effects of mutations on the structural integrity of the bromodomain using the ¹⁵N-HSQC spectra. The amide ¹H/¹⁵N resonances of the mutant proteins were compared to those of the wild-type bromodomain to determine whether the particular mutations led to global or local structure disruption. Severe line-broadening of the amide resonances indicates protein conformational exchange due to decreased structural stability resulting from point mutations. Structural integrity of the mutant proteins is expressed here relative to that of the wild type: +, as stable as wild type; ++, mildly destabilized; +++, moderately destabilized; +, completely unfolded.

† The ligand-binding affinity (K_D) of the bromodomain proteins was estimated by the following chemical shift changes of amide peaks in the ¹⁵N-HSQC spectra as a function of the ligand concentration.

‡ There was no detectable ligand binding in the NMR titration.

§ Ligand-binding affinity was significantly reduced and beyond the limit for reliable measurements by NMR titration.

In conclusion, our results indicate that the bromodomains are acetyl-lysine-binding domains, which, to our knowledge, makes them the first protein modules to exhibit such interactions. Like other modular domains, such as Src homology-2 (SH2) and phosphotyrosine-binding (PTB) domains¹⁹, that specifically interact with phosphotyrosine-containing proteins, the bromodomain/acetyl-lysine recognition could regulate protein–protein interactions by lysine acetylation. Our finding may explain why the bromodomain is indispensable for the function of GCN5 in yeast²⁰, and supports the hypothesis that bromodomains contribute to highly specific histone acetylation by tethering transcriptional HATs to specific chromosomal sites¹. Bromodomain–AcK binding could also be important for the assembly and activity of multi-protein complexes in transcriptional activation. Our results should help in the determination of specific molecular mechanisms used by the family of bromodomains in chromatin remodelling and transcriptional activation. □

Methods

Sample preparation. The bromodomain of P/CAF (residues 719–832) was subcloned into the pET14b expression vector (Novagen) and expressed in *Escherichia coli* BL21(DE3) cells. We prepared uniformly ¹⁵N- and ¹⁵N/¹³C-labelled proteins by growing bacteria in a minimal medium containing ¹⁵NH₄Cl with or without ¹³C₆-glucose. A uniformly ¹⁵N/¹³C-labelled and fractionally deuterated protein sample was prepared by growing the cells in 75% ²H₂O. The bromodomain was purified by affinity chromatography on a nickel–IDA column (Invitrogen), followed by the removal of the poly-His tag by thrombin cleavage. The final purification of the protein was achieved by size-exclusion chromatography. We prepared acetyl-lysine-containing peptides on a MiliGen 9050 peptide synthesiser (Perkin Elmer) using Fmoc/HBTU chemistry. Acetyl-lysine was incorporated using the reagent Fmoc-Ac-Lys with HBTU/DIPEA activation. NMR samples contained ~1 mM protein in 100 mM phosphate buffer, pH 6.5, and 5 mM perdeuterated DTT and 0.5 mM EDTA in H₂O/²H₂O (9/1) or ²H₂O.

NMR spectroscopy. We acquired NMR spectra at 30 °C on a Bruker DRX600 or DRX500 spectrometer. The backbone ¹H, ¹³C and ¹⁵N resonances were assigned by using deuterium-decoupled triple-resonance experiments of HNCACB and HN(CO)CACB²¹, recorded with the uniformly ¹⁵N/¹³C-labelled and fractionally deuterated protein. The side-chain atoms were assigned from 3D HCCH-TOCSY²² and (H)C(CO)NH-TOCSY²³ data collected on the uniformly ¹⁵N/¹³C-labelled protein. Stereospecific assignments of the methyl groups of valine and leucine residues were obtained using a fractionally ¹³C-labelled sample²⁴. The NOE-derived distance restraints were obtained from ¹⁵N- or ¹³C-edited 3D NOESY spectra²². We determined ϕ -angle restraints based on the ³J_{H_N,H α} coupling constants measured in a 3D HNHA spectrum²². Slowly exchanging amide protons were identified from a series of 2D ¹⁵N-HSQC spectra recorded after the H₂O buffer was changed to a ²H₂O buffer. The intermolecular NOEs used in defining the structure of the bromodomain/acetyl-histamine complex were detected in ¹³C-edited (F_1), ¹³C/¹⁵N-filtered (F_3) 3D NOESY spectra²². All NMR spectra were processed with the NMRPipe/NMRDraw programs and analysed using NMRView²⁵.

Structure calculations. Structures of the bromodomain were calculated with a distance geometry/simulated annealing protocol using the X-PLOR program²⁶. A total of 1,324 manually assigned NOE-derived distance restraints were obtained from the ¹⁵N- and ¹³C-edited NOE spectra. We further analysed the NOE spectra by the iterative automated assignment procedure using ARIA²⁷, which integrates with X-PLOR for structure calculations. A total of 1,519 unambiguous and 590 ambiguous distance restraints were identified from the NOE data by ARIA, many of which were checked and confirmed manually. The ARIA-assigned distance restraints agreed with the structures calculated by using only the manually assigned NOE distance restraints, 28 hydrogen-bond distance restraints for 14 hydrogen bonds, and 54 ϕ -angle restraints. The final structure calculations employed a total of 3,515 NMR experimental restraints obtained from the manual and the ARIA-assisted assignments, 2,843 of which were unambiguously assigned NOE-derived distance restraints (1,077 intra-residue, 621 sequential, 550 medium-range and 595 long-range NOEs). For the ensemble of the final 30 structures, no distance and torsional angle restraints

were violated by more than $0.3\text{ \AA}$ and 5° , respectively. The total distance violation, and dihedral violation energies were 178.7 ± 2.4 , 41.6 ± 0.9 and $0.50 \pm 0.06\text{ kcal mol}^{-1}$, respectively. The Lennard-Jones potential, which was not used during any refinement stage, was $-526.2 \pm 16.8\text{ kcal mol}^{-1}$ for the final structures. Ramachandran plot analysis of the final structures (residues 727–828) with Procheck-NMR²⁸ showed that $71.0 \pm 0.6\%$, $23.8 \pm 0.6\%$, $3.5 \pm 0.2\%$ and $1.7 \pm 0.2\%$ of the non-glycine and non-proline residues were in the most favourable, additionally allowed, generously allowed and disallowed regions, respectively. The corresponding values for the residues in the four α -helices (residues 727–743, 770–776, 785–802 and 807–827) were $88.9 \pm 0.4\%$, $11.0 \pm 0.4\%$, $0.1 \pm 0.1\%$ and $0.0 \pm 0.0\%$, respectively. The structure of the bromodomain/acetyl-histamine complex was determined using the free-form structure and an additional 25 intermolecular and 5 intra-ligand NOE-derived distance restraints.

Site-direction mutagenesis. We prepared mutant proteins using the QuikChange site-directed mutagenesis kit (Stratagene). The presence of appropriate mutations was confirmed by DNA sequencing.

Ligand titration. We performed ligand titration by recording a series of 2D ^{15}N - and ^{13}C -HSQC spectra on the uniformly ^{15}N - and $^{15}\text{N}/^{13}\text{C}$ -labelled bromodomain ($\sim 0.3\text{ mM}$), respectively, in the presence of different amounts of ligand concentration ranging from 0 to $\sim 2.0\text{ mM}$. The protein sample and the stock solutions of the ligands were all prepared in the same aqueous buffer containing 100 mM phosphate and 5 mM perdeuterated DTT at pH 6.5.

Received 12 January; accepted 8 April 1999.

1. Brownell, J. E. & Allis, C. D. Special HATs for special occasions: Linking histone acetylation to chromatin assembly and gene activation. *Curr. Opin. Genet. Dev.* **6**, 176–184 (1996).
2. Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349–352 (1997).
3. Wolffe, A. P. Sinful repression. *Nature* **387**, 16–17 (1997).
4. Shikama, N., Lyon, J. & Thangue, N. B. L. The p300/CBP family: Integrating signals with transcription factors and chromatin. *Trends Cell Biol.* **7**, 230–236 (1997).
5. Haynes, S. R. *et al.* The bromodomain: A conserved sequence found in human, *Drosophila* and yeast proteins. *Nucleic Acids Res.* **20**, 2603–2603 (1992).
6. Jeanmougin, F., Wurtz, J. M., Douarin, B. L., Chambon, P. & Losson, R. The bromodomain revisited. *Trends Biochem. Sci.* **22**, 151–153 (1997).
7. Brownell, J. E. *et al.* Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84**, 843–851 (1996).
8. Ogryzko, V. V., Schiltz, O. L., Russanova, V., Howard, B. H. & Nakatani, Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* **87**, 953–959 (1996).
9. Bannister, A. J. & Kouzarides, T. The CBP co-activator is a histone acetyltransferase. *Nature* **384**, 641–643 (1996).
10. Yang, X.-J., Ogryzko, V. V., Nishikawa, J.-I., Howard, B. H. & Nakatani, Y. A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature* **382**, 319–324 (1996).
11. Puri, P. L. *et al.* Differential roles of p300 and PCAF acetyltransferases in muscle differentiation. *Cell* **1**, 35–45 (1997).
12. Richardson, J. S. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34**, 167–339 (1981).
13. Presnell, S. R. & Cohen, F. E. Topological distribution of four- α -helix bundles. *Proc. Natl Acad. Sci. USA* **86**, 6592–6596 (1989).
14. Weber, P. C. & Salemme, F. R. Structural and functional diversity in 4- α -helical proteins. *Nature* **287**, 82–84 (1980).
15. Kurokawa, R. *et al.* Differential use of CREB binding protein-coactivator complex. *Nature* **279**, 700–703 (1998).
16. Chen, H. *et al.* Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300. *Cell* **90**, 569–580 (1997).
17. Kuo, M.-H. *et al.* Transcription-linked acetylation by Gcn5p of histone H3 and H4 at specific lysines. *Nature* **383**, 269–272 (1996).
18. Dutnall, R. N., Tafrov, S. T., Sternglanz, R. & Ramakrishnan, V. Structure of the histone acetyltransferase Hat1: A paradigm for the GCN5-related N-acetyltransferase superfamily. *Cell* **94**, 427–438 (1998).
19. Pawson, T. Protein modules and signalling networks. *Nature* **373**, 573–580 (1995).
20. Geogakopoulos, T., Gounalaki, N. & Thireos, G. Genetic evidence for the interaction of the yeast transcriptional co-activator proteins GCN5 and ADA2. *Mol. Gen. Genet.* **246**, 723–728 (1995).
21. Yamazaki, T., Lee, W., Arrowsmith, C. H., Mahandirani, D. R. & Kay, L. E. A suite of triple resonance NMR experiments for the backbone assignment of ^{15}N , ^{13}C , ^2H labeled proteins with high sensitivity. *J. Am. Chem. Soc.* **116**, 11655–11666 (1994).
22. Clore, G. M. & Gronenborn, A. M. Multidimensional heteronuclear nuclear magnetic resonance of proteins. *Meth. Enzymol.* **239**, 249–363 (1994).
23. Logan, T. M., Olejniczak, E. T., Xu, R. X. & Fesik, S. W. A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. *J. Biomol. NMR* **3**, 225–231 (1993).
24. Neri, D., Szyperski, T., Otting, G., Senn, H. & Wüthrich, K. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. *Biochemistry* **28**, 7510–7516 (1989).
25. Johnson, B. A. & Blevins, R. A. NMRView: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
26. Brünger, A. T. *X-PLOR Version 3.1: A System for X-Ray Crystallography and NMR* (Yale University Press, New Haven, 1993).
27. Nilges, M. & O'Donoghue, S. Ambiguous NOEs and automated NOE assignment. *Progr. NMR Spectr.* **32**, 107–139 (1998).
28. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).

29. Carson, M. Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 958–961 (1991).

30. Nicholls, A., Bharadwaj, R. & Honig, B. GRASP: Graphical representation and analysis of surface properties. *Biophys. J.* **64**, 166–170 (1993).

Acknowledgements. We thank M. Sattler, M. Nilges, M. Rosen, R. P. Meadows, and C. Escalante for technical advice; O. Plotnikova for assistance in the preparation of mutant proteins; I. Wolf for peptide synthesis; and D. Logothetis, H. Weinstein, L. Shapiro, R. Margolske and A. Farooq for their suggestions and for critical reading of the manuscript. This work was supported by discretionary funds from the Mount Sinai School of Medicine (to M.M.Z.) and NIH grants (to A.A.).

Correspondence and requests for materials should be addressed to M.M.Z. (e-mail: zhoum@inka.mssm.edu). Coordinates for the NMR structure of the P/CAF bromodomain have been deposited in the Brookhaven Protein Data Bank under accession code 1B91.

Enzyme dynamics and hydrogen tunnelling in a thermophilic alcohol dehydrogenase

Amnon Kohen*, Raffaele Cannio†, Simonetta Bartolucci‡ & Judith P. Klinman*§

* Department of Chemistry and § Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720, USA

† ISA CNR, via Roma, 83100 Avellino, Italy

‡ Dipartimento di Chimica Organica e Biologica, Università di Napoli “Rederico II”, via Mezzocannone 16, 80134 Napoli, Italy

Biological catalysts (enzymes) speed up reactions by many orders of magnitude using fundamental physical processes to increase chemical reactivity. Hydrogen tunnelling has increasingly been found to contribute to enzyme reactions at room temperature¹. Tunnelling is the phenomenon by which a particle transfers through a reaction barrier as a result of its wave-like property^{1–3}. In reactions involving small molecules, the relative importance of tunnelling increases as the temperature is reduced⁴. We have now investigated whether hydrogen tunnelling occurs at elevated temperatures in a biological system that functions physiologically under such conditions. Using a thermophilic alcohol dehydrogenase (ADH), we find that hydrogen tunnelling makes a significant contribution at 65°C ; this is analogous to previous findings with mesophilic ADH at 25°C (ref. 5). Contrary to predictions for tunnelling through a rigid barrier, the tunnelling with the thermophilic ADH decreases at and below room temperature. These findings provide experimental evidence for a role of thermally excited enzyme fluctuations in modulating enzyme-catalysed bond cleavage.

Structural studies of enzymes have influenced our view of catalysis enormously by providing three-dimensional pictures of active sites. But although these time-averaged structures are extremely useful as a starting point for evaluating and modelling catalysis, they generally fail to incorporate the wide range of dynamic motion that can occur in proteins⁶. It is a major experimental challenge to examine the relationship between the dynamics of a protein and its activity. Various studies have shown a relationship between protein dynamics and the physical steps of the enzymatic process, that is, those steps involving the formation and breakdown of catalytic complexes^{7,8}, but a far greater problem is whether protein dynamics have a direct role in the catalysis of bond forming and cleaving. Our present findings on hydrogen transfer under physiological conditions cannot be explained without invoking both quantum mechanics and enzyme dynamics.

Recent literature has demonstrated the importance of quantum-mechanical tunnelling in enzyme-catalysed hydrogen-transfer reactions¹. Indirect evidence that dynamical motion is involved comes from the demonstration that surface glycosylation of glucose oxidase decreases the extent of quantum behaviour observed in

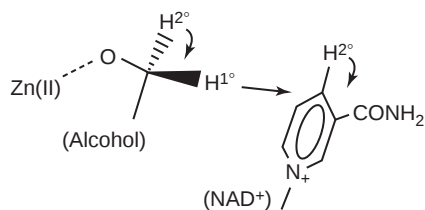


Figure 1 Motion of the primary (1°) and secondary (2°) hydrogens in the reaction of alcohol dehydrogenase.

enzyme-catalysed hydrogen transfer⁹. The present study investigated hydrogen tunnelling in an alcohol dehydrogenase from *Bacillus stearothermophilus* LLD-R strain, a thermophilic protein (referred to as ADH-hT) that normally functions at 65 °C (refs 10,11). ADH-hT is very similar to the more extensively studied mesophilic ADHs (from horst liver and yeast), having 55% sequence identity, including conservation of most active-site residues, and containing the structural and catalytic zinc ions found in mesophilic ADHs. With thermophilic enzymes, the study of catalysis as a function of temperature has often demonstrated biphasic Arrhenius behaviour, with an increased enthalpy of activation at temperatures below the break-point in plots of $\ln(k)$ against the reciprocal of the absolute temperature ($1/T$) (refs 12–15). This behaviour has been attributed to increased protein rigidity at reduced temperatures.

Our study of hydrogen tunnelling as a function of temperature using thermophilic ADH-hT allows us to examine how protein mobility affects the contribution of hydrogen tunnelling to the reaction rate. Specifically, if protein mobility is a prerequisite for effective tunnelling, a decreased tunnelling contribution is predicted as the temperature is reduced below the optimal temperature for ADH-hT. This is in direct contrast to models of hydrogen tunnelling through a rigid barrier, which predict an increased contribution of tunnelling to rate with decreased temperature^{3,4}.

We measured tunnelling with ADH-hT according to protocols previously established for demonstrating tunnelling with mesophilic enzymes^{1,5}. One of these protocols involves the competitive measurement of multiple kinetic isotope effects (KIEs, that is, k_D/k_T and k_H/k_T where k_H , k_D and k_T are rate constants for reaction of protium, deuterium and tritium. It is expected that $(k_D/k_T)^{3.3} = k_H/k_T$ under conditions of semi-classical behaviour¹⁶, that is, when tunnelling makes no contribution to reaction rate. Measurements conducted under conditions where the hydrogen-transfer step is only partially rate limiting deflate the exponent to below 3.3, while enhanced tunnelling of protium in relation to deuterium can lead to an exponent larger than 3.3 (refs 16–18). Isotope effects can be measured in either the primary (1°) or the secondary (2°) position (see Fig. 1). Exponents larger than 3.3 are most readily detected in secondary KIE measurements under conditions of both tunnelling and coupled motion between the primary and secondary hydrogens of the substrate^{16–18} (Fig. 1).

Figure 2a–d shows the temperature dependencies measured for the primary and secondary k_H/k_T and k_D/k_T isotope effects with ADH-hT, and indicates that the KIEs are essentially temperature independent between 65 °C and 30 °C but rise steeply with temperature below 30 °C. The exponents relating H/T and D/T $(k_D/k_T)^{\text{exponent}} = k_H/k_T$ for secondary KIEs (Fig. 2f) are much higher than the semi-classical limit between 65 °C to 30 °C and decrease towards semi-classical values as the temperature is reduced from 30 °C to 5 °C. The data in Fig. 2 provide clear evidence for significant hydrogen tunnelling at elevated temperature with ADH-hT (at its physiological temperature). The decrease in the secondary exponents at temperatures below 30 °C cannot be due to a change in the rate-determining step that makes the hydrogen-transfer step less rate determining, because primary exponents are not suppressed

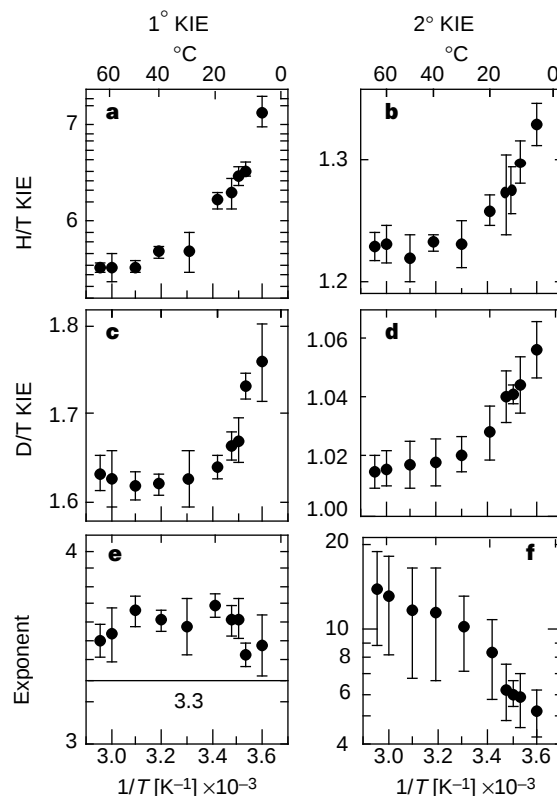


Figure 2 Competitive primary (1°) and secondary (2°) kinetic isotope effects (KIEs) on logarithmic abscissa versus $1/T$, ranging from 5 °C to 65 °C. **a**, 1° H/T KIEs and **b**, 2° H/T KIEs were measured simultaneously for the ADH-hT-catalysed oxidation of benzyl alcohol by NAD^+ . **c**, 1° D/T KIEs and **d**, 2° D/T KIEs were measured likewise. **e**, **f**, The exponent that relates 1° and 2° H/T and D/T measurements. The solid line in **e** indicates the semi-classical upper limit of 3.3. The relationship of H/T to D/T KIE is calculated from: $k_H/k_T = (k_D/k_T)^{\text{exponent}}$. Standard errors (error bars) are calculated from:

$$\Delta \text{exponent} = \sqrt{\left(\frac{\Delta(k_H/k_T)}{k_H/k_T} / \ln(k_D/k_T) \right)^2 + \left(\frac{\Delta(k_D/k_T)}{k_D/k_T} \cdot \frac{\ln(k_H/k_T)}{(\ln(k_D/k_T))^2} \right)^2}$$

below 3.3 (Fig. 2e), and the measured primary KIEs become larger as the temperature is reduced below 30 °C (Fig. 2a, c). Because rigid models predict that tunnelling will be more significant at reduced temperatures, Fig. 2 indicates a temperature-dependent transition in the protein's behaviour that impairs hydrogen tunnelling.

A second major tool for investigating hydrogen tunnelling comes from comparing Arrhenius plots for the reaction of the isotopes of hydrogen (Fig. 3). It is well accepted that for reactions with no tunnelling contribution, the intercept (A) of Arrhenius plots for L/T KIEs, denoted as A_L/A_T (where L is H or D) will be close to unity (for example, Fig. 3, region I)². By contrast, when protium tunnelling is greater than with deuterium or tritium, the experimentally measured KIEs will result in inverse values for A_L/A_T (for example, Fig. 3, region II)². Conversely, when tunnelling through a rigid barrier dominates a reaction, slopes of Arrhenius plots for all isotopes approach zero and their A_L/A_T values are significantly greater than unity (Fig. 3, region III)⁴. Analysis of the primary KIE data in Fig. 2 allows an estimate of both A_H/A_T and A_D/A_T (see Table 1). Once again, these data implicate more tunnelling at the higher temperature range (65 °C to 30 °C, $A_L/A_T > 1$) than between 30 °C and 5 °C ($A_L/A_T < 1$).

Our finding of values for $A_L/A_T > 1$ between 65 °C and 30 °C is particularly intriguing in the context of the mechanism by which ADH-hT facilitates tunnelling. If the behaviour of ADH-hT is analogous to regime III of Fig. 3 (tunnelling through a rigid barrier),

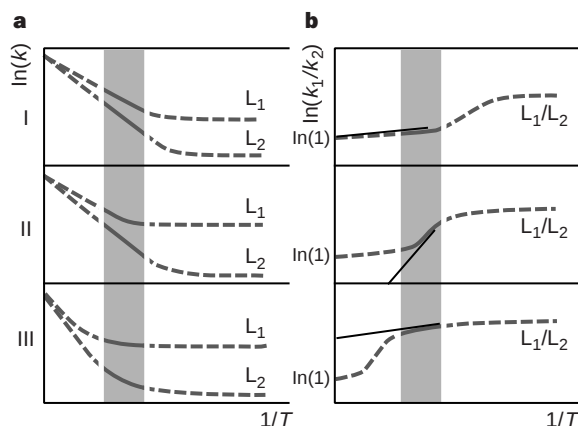


Figure 3 **a**, Arrhenius plot of $\ln(k)$ against $1/T$ for a light isotope (L_1) and a heavier isotope (L_2). **b**, Arrhenius plot of $\ln(k_1/k_2)$ against $1/T$. The experimental temperature range is highlighted in grey. The L_1/L_2 KIE on the Arrhenius pre-exponential factors (A_1/A_2) is the intercept of the tangent to the curve at the experimental temperature range (dark lines). Three experimental systems are presented: system I, with no tunnelling contribution, A_1/A_2 of unity; system II, with some tunnelling contribution to the L_1 reaction whose plot curves upward, $A_1/A_2 < 1$; system III, where both isotopes tunnel significantly, $A_1/A_2 > 1$. (Reproduced with some modification from ref. 1.)

the experimental enthalpy of activation ($\Delta H^\ddagger$) for C–L-bond cleavage would be expected to be close to zero and KIEs would be enormous⁴. In contrast, if tunnelling occurs by a vibrationally enhanced mechanism, values measured for the individual $\Delta H^\ddagger$ may be large and the magnitude of isotope effects could fall in the semi-classical range¹⁹. To investigate this distinction, we analysed non-competitive isotope effects by comparing the rate of oxidation of $[1-^1\text{H}_2]$ -alcohol by ADH-hT with that of $[1-^2\text{H}_2]$ -alcohol.

We measured initial velocities while varying both alcohol and nicotinamide adenine dinucleotide (NAD^+) concentrations, thus allowing the rate constant when both substrates are at saturating concentrations (k_{cat}), to be estimated as a function of temperature. The resulting Arrhenius plots (Fig. 4) have slopes at elevated temperatures that are both large and essentially identical for protonated and deuterated substrates, with $A_{\text{H}}/A_{\text{D}} > 1$ (Table 2). Below the transition temperature of about 30 °C, the slopes for protonated and deuterated substrates begin to diverge, with increased $\Delta H^\ddagger$ values and $A_{\text{H}}/A_{\text{D}} < 1$. These non-competitive rate data lead to several key confirmatory findings. First, the fact that the curves for protonated and deuterated substrates break at similar temperatures argues against a change in the rate-determining step, once again implicating a physical change in the protein as the origin of the observed behaviour. This is supported by the finding that the size of the KIE increases below 30 °C. Additionally, the non-competitive KIEs on k_{cat} and $k_{\text{cat}}/K_{\text{m}}$ are similar (data not shown), consistent with a random equilibrium kinetic mechanism in which C–H-bond cleavage is close to fully rate-determining across the full temperature range. As was concluded from the competitive KIEs described above, the different $\Delta H^\ddagger$ values for protonated alcohol and deuterated alcohol below 30 °C indicate greater semi-classical behaviour, that is, a reduction in tunnelling as the temperature is lowered.

Table 1 $A_{\text{L}}/A_{\text{T}}$ from primary KIEs measured under competitive conditions

Temperature range	$A_{\text{H}}/A_{\text{T}}$	$A_{\text{D}}/A_{\text{T}}$
5–30 °C	0.26 (± 0.23)	0.23 (± 0.14)
Semi-classical*	0.60 to 1.67	0.90 to 1.22
30–65 °C	4.3 (± 0.6)	1.73 (± 0.26)

$A_{\text{L}}/A_{\text{T}}$ are isotope effects on Arrhenius pre-exponential factors where L is H or D (see text and Fig. 2).

* Refs 2, 19.

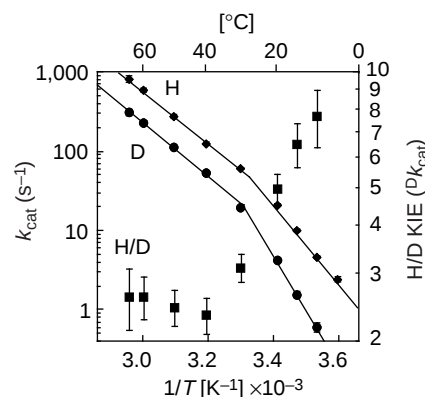


Figure 4 Arrhenius plots of k_{cat} for protonated (diamonds) and deuterated (circles) substrates (left abscissa), and Arrhenius plot of their KIE (squares) (right abscissa).

Our results show that tunnelling is very significant at elevated temperatures under conditions where the thermophilic enzyme is optimally functional. The extent of tunnelling decreases with temperature, opposite to the trend predicted from simple chemical models and measurements, and indicating a transition for the protein at approximately 30 °C. In light of the substantial evidence for an increased rigidity of thermophilic enzymes with reduced temperature^{12,13,15,20}, we ascribe the decrease in tunnelling at reduced temperatures to reduced protein mobility required to facilitate tunnelling. This model of vibrationally enhanced tunnelling is consistent with the high-temperature data in which $\Delta\Delta H^\ddagger$ for protonated compared with deuterated substrate is close to zero while the individual values for $\Delta H^\ddagger$ remain sizeable. Although it has been postulated theoretically that enzymes work by dynamically enhanced mechanisms (see below), the present results provide the first experimental evidence to support this.

Since the work of Vol'kenshtein²¹, increasingly sophisticated theories have been advanced describing a possible link between protein motion and hydrogen tunnelling^{22–26}. Work by Borgis and Hynes²⁵, Bruno and Bialek^{26,27,29} and most recently Antoniou and Schwartz^{23–27} have stressed the importance of orthogonal coordinates involving a temperature-dependent, isotope-independent variation in the relative energy levels for reactant and product, and a temperature and isotope-dependent variation of the internuclear distance. A key feature of all these models is that the temperature dependence for the tunnelling reaction derives from the reaction exothermicity and the energetic terms for protein rearrangements, that is, $\Delta H^\ddagger$ need not reflect the barrier height for C–H-cleavage as formulated according to classical transition-state theory³. Although a change in the average protein conformation cannot be ruled out, the activity transition for ADH-hT can be understood in the context of a motion(s) that decreases below 30 °C. At reduced temperatures the amplitude of the promoting vibration(s) for ADH-hT decreases below a level required for effective tunnelling. The reaction begins to approximate more classical behaviour with less tunnelling, an increased enthalpy of activation and a rate that is an order of magnitude less than that predicted from the high-temperature behaviour.

Table 2 Parameters derived from non-competitive measurements

Parameter	Temperature range (°C)	
	5–30	30–65
$\Delta H_{30^\circ\text{C}}^\ddagger$ (H) (kcal mol ⁻¹)	23.6 (± 0.6)	14.6 (± 0.3)
$\Delta H_{30^\circ\text{C}}^\ddagger$ (D) (kcal mol ⁻¹)	31.4 (± 1.7)	15.1 (± 0.5)
$\Delta H^\ddagger(\text{D}) - \Delta H^\ddagger(\text{H})$ (kcal mol ⁻¹)	7.8 (± 1.8)	0.5 (± 0.6)
$A_{\text{H}}/A_{\text{D}}$	$1 \times 10^{-5} (\pm 1 \times 10^{-5})$	
		2.2 (± 1.1)

See also Fig. 4.

Our findings have important implications for the catalytic behaviour of enzymes. The comparison of thermophilic and mesophilic species (25 °C for the yeast ADH, 65 °C for ADH-hT) verifies the old notion that different enzymes impose similar environments on their active sites at their physiological temperatures^{13,28}. Our comparison of the tunnelling behaviour for the thermophilic enzyme across a sizeable temperature range (65–5 °C) suggests a role for protein mobility in C–H-bond cleavage. It is perhaps not surprising that, for C–H-bond cleavage reactions, a partitioning of thermal energy into low-frequency protein modes is a major factor. These modes can become significantly excited at ambient temperature, in contrast to the C–H stretching mode, which is expected to be largely confined to its zero-point energy level. The implication of protein dynamics in modulating chemical steps makes it easier to understand why biomimetic models, as well as catalytic antibodies, often fall short of the rate accelerations observed with enzymes. The idea of a dynamic coupling of protein vibrational modes into the chemical reaction coordinate should increasingly aid our understanding of biological catalysis.

Note added in proof: Studies by Basram *et al.* (*Biochemistry* **38**, 3218–3222 (1999)) on the hydrogen transfer catalysed by a mesophilic methylamine dehydrogenase show isotope effects and enthalpies of activation consistent with a role for vibrationally enhanced tunnelling. □

Methods

Enzyme preparation. ADH-hT from *B. stearothermophilus* LLD-R strain was expressed and purified as described before^{10,11} with minor modification.

Competitive KIEs. All competitive KIEs reported were measured under Ar with 4 mM NAD⁺, 50 mM potassium phosphate, pH 7.0, 300 mM semicarbazide hydrochloride, and 1.5 mM dithiothreitol, between 5 and 65 °C. The temperature was thermostated to ± 0.1 °C in a Neslab water bath, and the pH was adjusted at the experimental temperature. The kinetic experiments were conducted and analysed as described previously⁵, with minor modifications. Each reported KIE is an average of 14–35 points measured in several different experiments.

Non-competitive measurements. Initial velocities were measured over a temperature range of 5–65 °C in 50 mM potassium phosphate, pH 7.0, by following the 340-nm absorption of the product NADH using [1, 1-¹H₂] or [1, 1-²H₂] benzyl alcohol. NAD⁺ concentrations were in the range 0.2–4.0 mM and alcohol concentrations were 0.75–20 mM. Cofactor concentration above 10 mM and alcohol concentration above 50 mM showed some substrate inhibition. For each k_{cat} measurement, five substrate concentrations were varied with five cofactor concentrations and the initial velocities were fitted to the equation: $v = k_{\text{cat}}[E][A][B]/(K_{\text{ia}}[B] + K_A[B] + K_B[A] + [A][B])$, where [E] is the enzyme concentration, [A] and [B] are the cofactor and substrate concentrations, respectively, K_A and K_B are their Michaelis constants and K_{ia} is the cofactor dissociation constant. The slope and intercept of the Arrhenius

plots at the different temperature ranges were obtained from the error-weighted nonlinear fit. The ratio of the values of the protonated versus the deuterated substrates yielded the H/D KIEs.

Received 26 October 1998; accepted 16 March 1999.

- Kohen, A. & Klinman, P. J. Enzyme catalysis: beyond classical paradigms. *Acc. Chem. Res.* **31**, 397–404 (1998).
- Bell, R. P. *The Tunneling Effect in Chemistry* (Chapman & Hall, London, 1980).
- Steinfeld, J. I., Francisco, J. S. & Hase, W. L. (eds) *Chemical Kinetics and Dynamics* (Prentice Hall, Upper Saddle River, NJ, 1998).
- Benderskii, V. A., Makarov, D. E. & Wight, C. A. *Adv. Chem. Phys.* **88**, 151–207 (1994).
- Chai, Y., Murray, C. J. & Klinman, J. P. Hydrogen tunneling in enzyme reactions. *Science* **243**, 1325–1330 (1989).
- Brooks, C. L. I., Karplus, M. & Pettitt, B. M. *Adv. Chem. Phys.* **71**, 14–23; 33–59; 75–233; (1988).
- Rasmussen, B. F., Stock, A. M., Ringe, D. & Petsko, G. A. Crystalline ribonuclease-A loses function below the dynamical transition at 220-K. *Nature* **357**, 423–424 (1992).
- Rudd, P. M. *et al.* Glycoforms modify the dynamic stability and functional activity of an enzyme. *Biochemistry* **33**, 17–22 (1994).
- Kohen, A., Jonsson, T. & Klinman, P. J. Effect of protein glycosylation on catalysis: changes in hydrogen tunneling and enthalpy of activation in the glucose oxidase reaction. *Biochemistry* **36**, 2603–2611 (1997).
- Cannio, R., Rossi, M. & Bartolucci, S. A few amino acid substitutions are responsible for the higher thermostability of a novel NAD(+)–dependent bacillar alcohol dehydrogenase. *Eur. J. Biochem.* **22**, 345–352 (1994).
- Guagliardi, A. *et al.* Purification and characterization of the alcohol dehydrogenase from a novel strain of *Bacillus stearothermophilus* growing at 70 °C. *Int. J. Biochem. Cell Biol.* **28**, 239–246 (1996).
- Lakatos, S., Halasz, G. & Zavodszky, P. Conformational stability of lactate dehydrogenase from *Bacillus thermus-aquaticus*. *Biochem. Soc. Trans.* **6**, 1195–1197 (1978).
- Zavodszky, P., Kardos, J., Svingor, A. & Petsko, G. A. Adjustment of conformational flexibility is a key event in the thermal adaptation of protein. *Proc. Natl Acad. Sci. USA* **95**, 7406–7411 (1998).
- Varley, P. G. & Pain, R. H. Relation between stability, dynamics and enzyme activity in 3-phosphoglycerate kinases from yeast and *Thermus thermophilus*. *J. Mol. Biol.* **220**, 531–538 (1991).
- Wrba, A., Schweiger, A., Schultes, V., Jaenicke, R. & Zavodszky, P. Extremely thermostable D-glyceraldehyde-3-phosphate dehydrogenase from the eubacterium *Thermotoga maritima*. *Biochemistry* **29**, 7584–7592 (1990).
- Lin, S. & Saunders, W. H. Tunneling in elimination reactions—structural effects on the secondary beta-tritium isotope effects. *J. Am. Chem. Soc.* **116**, 6107–6110 (1994).
- Saunders, W. H. Calculations of isotope effects in elimination reactions. New experimental criteria for tunneling in slow proton transfers. *J. Am. Chem. Soc.* **107**, 164–169 (1985).
- Rucker, J. & Klinman, J. P. Computational study of tunneling and coupled motion. *J. Am. Chem. Soc.* **121**, 1997–2006 (1999).
- Melander, L. & Saunders, W. H. *Reaction Rates of Isotopic Molecules* (Krieger, Malabar, FL, 1987).
- Vihinen, M. Relationship of protein flexibility to thermostability. *Protein Eng.* **1**, 477–480 (1987).
- Vol'kenshtein, M. V., Dogonadze, R. R., Madumarov, A. K., Urushadze, Z. D. & Kharkats, Y. I. The theory of enzyme catalysis. *Mol. Biol.* **6**, 347–353 (1972).
- Hwang, J. K. & Warshel, A. On the relationship between the dispersed polaron and spin-boson models. *Chem. Phys. Lett.* **271**, 223–225 (1997).
- Antonioni, D. & Schwartz, S. D. Large kinetic isotope effect in enzymatic proton transfer and the role of substrate oscillations. *Proc. Natl Acad. Sci. USA* **94**, 12360–12365 (1997).
- Borgis, D. & Hynes, J. T. In *The Enzyme Catalysis Process* (eds Cooper, A., Houben, J. & Chien, L.) 293–303 (Plenum, New York, 1989).
- Bruno, W. J. & Bialek, W. Vibrationally enhanced tunneling as a mechanism for enzymatic hydrogen transfer. *Biophys. J.* **63**, 689–699 (1992).
- Borgis, D. & Hynes, J. T. Dynamical theory of proton tunneling transfer rates in solution—general formulation. *Chem. Phys.* **170**, 315–346 (1993).
- Antonioni, D. & Schwartz, S. D. Activated chemistry in the presence of a strongly symmetrically coupled vibration. *J. Chem. Phys.* **108**, 3620–3625 (1998).
- Somero, G. N. & Siebenaller, J. F. Inefficient lactate dehydrogenases of deep-sea fishes. *Nature* **282**, 100–102 (1979).

Acknowledgements. This work was supported by the NSF (A.K. and J.P.K.) and EC-projects (R.C. and S.B.).

Correspondence and requests for materials should be addressed to J.P.K. (e-mail: klinman@socrates.berkeley.edu).

Of mice and mice

Two kits making it easier to see staining with mouse antibodies against a background of mouse tissue, and ELISA assays for insulin and FK506 feature in an immunologically inclined selection.

Hepatitis marker

From DPC

www.dpcweb.com

New assays for the Immulite 2000 automated immunoassay system

DPC has introduced three new assays for the identification of serological markers associated with hepatitis-B virus infection. Immulite 2000 assays are now available for the qualitative determination of HBsAg and Anti-HBc, and quantitative measurement of Anti-HBs. First marker to appear in serum is HBsAg, indicating an ongoing infection with HBV. Antihepatitis core antigen (Anti-HBc) appears next, peaking during the acute phase and remaining elevated in chronic infection. Antibodies to the surface antigen (anti-HBs) generally appear 6 months after infection and their presence represents recovery and immunity.

Reader Service No. 100

Nunc-Immuno Wash

From Nalge Nunc International

nunc.nalgenunc.com

A versatile wash for hand-held washing

The Nunc-Immuno Wash is available in 8- and 12-channel versions to fit 96-well Immuno plates and modules and in a 16-channel version for 384-well plates. It is suitable for labs performing a limited number of tests per trial or requiring an individual washing procedure for each application. The machine is connected to a vacuum source and a source of washing solution; the washing cannulae are placed in the wells of a row or column of a MicroWell plate or strip, and washing is initiated by the press of a button. The washer can be adjusted to all well shapes, including NucleoLink, by turning the two adjustment screws beside the cannulae.

Reader Service No. 101

Cell separation

From StemCell Technologies

www.stemcell.com

Add up the column inches

An 0.1-inch column has been added to the range of StemSep immunomagnetic cell separation devices for the processing of small samples. The optimal cell number per separation is 10^6 to 10^7 cells, with an extended range from 10^5 to 2×10^7 cells. Applications include the separation of clinical samples with limited cell numbers, the isolation of human cells from a single chimaeric mouse or the separation of cell populations from murine lymph nodes. At



Immulite 2000: up to 200 tests an hour.

the other end of the scale is the new 1-inch column for the isolation of rare cells where the large quantities must be processed to give a chance of finding the desired cell.

Reader Service No. 102

ActiSep

From Sterogene Bioseparations

www.sterogene.com

Elution medium to recover antibodies, but gently

ActiSep is a nondenaturing, neutral elution medium used to recover antibodies from immunoabsorbents and Protein A resins without the loss of biological activity. The ActiSep cocktail of components gently releases antibodies by individually breaking up the multiple interactions between the antibody and antigen/Protein A. ActiSep is non-toxic and all ActiSep components are compatible with pharmaceutical use. A Drug Master File is available in the United States and Europe.

Reader Service No. 103

TSA-Direct

From NEN

www.nenlifesci.com

Multi-colour fluorescence detection in immunohistochemistry and in situ hybridization

Two new TSA (tyramide signal amplification) kits from NEN Life Science Products combine the high fluorescence of cyanine dyes with NEN's proprietary TSA technology. TSA-Direct (cyanine 5) is intended for immunohistochemistry, and TSA-Direct (cyanine 5 FISH) for *in situ* hybridization. The new kits join existing TSA kits contain-

ing other dyes, including cyanine 3-tyramide, to offer multi-colour detection capabilities in multiple-labelling applications. As well as enabling the detection of very low levels of target proteins and nucleic acids, NEN claim that TSA reduces the concentrations of primary or secondary antibodies required, thus saving valuable reagents.

Reader Service No. 104

EnVision+

From Dako

www.dako.com

Biotin-free immunostaining

In EnVision+ detection systems, labelled polymer is used to give rapid, biotin-free immunostaining. The process is based on an HRP-labelled dextran polymer to which secondary antibodies are conjugated. By incorporating link and label onto the polymer backbone a two-step biotin-free assay is achieved. In addition, EnVision+ can be used for blot and ELISA applications. Reagents are supplied ready for use with mouse or rabbit primary antibodies on formalin-fixed paraffin-embedded tissues, cryostat sections and cell preps.

Reader Service No. 105

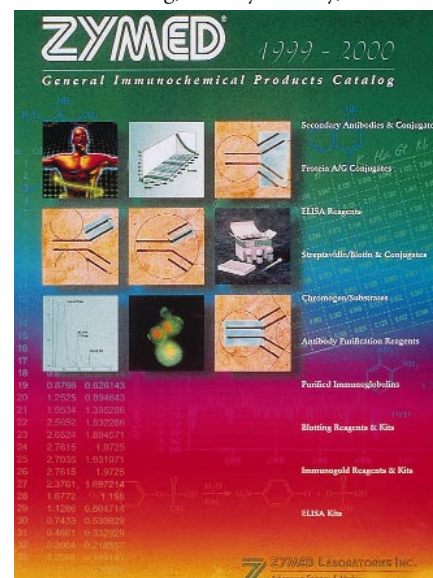
Immunochemicals catalogue

From Zymed

www.zymed.com

Antibodies and the rest

Zymed's 1999–2000 General Immunochemical Products catalogue contains more than 1,200 products for immunoassay systems including reagents for ELISA, western blotting, flow cytometry, immuno-



Y2K compliant: still 'live' come January.

precipitation and immunohistochemistry. Included are Zymed's affinity-purified anti-mouse, rat, rabbit and goat secondary antibodies. These are available unconjugated or conjugated to HRP, biotin, alkaline phosphatase, FITC, TRITC, Cy3, Cy5 and gold. Also available are isotype-specific antibodies to mouse IgG1, IgG2a, IgG2b, IgG3, IgA, and IgM; rat IgG1, IgG2a, IgG2b, IgG2c, IgA, IgE and IgM; and human IgG1, IgG2, IgG3 and IgG4.

Reader Service No. 106

Anti-PARP

From Kamiya Biomedical

www.kamiyabiomedical.com

A monoclonal to PARP

Poly (ADP-ribose) polymerase (PARP) is a eukaryotic protein involved in differentiation, DNA repair and the organization of chromatin. Kamiya has introduced a monoclonal antibody (clone C2-10, murine IgG1) raised against bovine PARP. This antibody reacts with PARP (116 kDa) from mouse, rat, hamster and primates, and with the 85-kDa CPP32/caspase 3 cleavage product. It does not react with PARP from chickens; reactivity to PARP from other species has not been tested. The antibody is suitable for immunocytochemistry, western blot and ELISA.

Reader Service No. 107

Leukocyte antigen clone

From BioGenex

www.biogenex.com

Clone LCA88, ready to use in immunostaining

CD45 (or LCA, leukocyte common antigen) is a transmembrane protein-tyrosine phosphatase abundant on the lymphocyte surface and reported to be involved in signalling through the T-cell antigen receptor. Monoclonal antibody LCA88 reacts specifically with CD45 in formalin-fixed, paraffin-embedded tissue sections. This antibody is available in concentrated form or as a ready-to-use reagent optimized for use with Super Sensitive detection systems. This antibody may be useful in the evaluation of malignant lymphoma and nonlymphoid tumours.

Reader Service No. 108

FK506 ELISA

From DiaSorin

www.diasorin.com

An immunoassay for the immunosuppressor FK506

DiaSorin's PRO-TRAC II FK506 ELISA is a competitive enzyme immunoassay based on a monoclonal antibody to FK506. This assay is run on a microtitre plate. Standards, controls and samples are extracted with an enzyme digestion reagent and added to the wells of the microtitre plate prior to addition



Bottle bank: the kit for FK506 assays.

of the anti-FK506 monoclonal. After a 30-min incubation at room temperature, the FK506-horseradish peroxidase conjugate is added and incubated for an additional 60 min. The wells are then washed and chromogen is added for a 15-min incubation. The reaction is stopped by addition of acid and the absorbance in each well is read at a dual wavelength of 450/630 nm.

Reader Service No. 109

ADVERTISEMENTS

STERNBERGER MONOCLONAL (SMI) ANTIBODIES FROM AFFINITI

The SMI series of antibodies to neurofilaments, tau, myelin basic protein, blood brain barrier (EBA), SNAP-25, and tubulins is available from AFFINITI.

We also offer many antibodies to proteins of neuroscientific interest, including GAD65/67, α , β and γ (NEW) Synucleins, heme oxygenase, neuronal nitric oxide synthase, PGP9.5 and β -catenin.

For data sheets or more information on AFFINITI's extensive range of products for neuroscience research please visit our website at:

[HTTP://WWW.AFFINITI-RES.COM](http://www.affiniti-res.com)

AFFINITI Research Products Ltd., Mamhead Castle, EXETER, EX6 8HD, UK
Tel. +44/0 1626 891010; Fax: +44/0 1626 891090; URL: <http://www.affiniti-res.com>

READER ENQUIRY NO. 106

The Mouse Kit™ for Gene Targeting in ES Cells

- ✓ Complete with embryonic stem (ES) cells, feeder cells and essential reagents (including FBS) for engineering one knockout mouse strain
- ✓ Low-passage AB2.2-Prime ES Cells® (129/SvEv)
- ✓ Uniform, high-quality ESQ feeder cells (mitotically inactivated STO cells) engineered for *neo* and *puro* resistance
- ✓ Prequalification of ESQ feeder cells and reagents for use with AB2.2-Prime ES Cells

*Manufactured by Lexicon Genetics Incorporated

Fax: (619) 535-0071
STRATAGENE
Tel: (619) 535-5400 or (800) 424-5444
www.stratagene.com

READER ENQUIRY NO. 109

Detect residual DNA with Wako's DNA Extractor Kit or Isolate Genomic DNA with DNA Extractor WB Kit
Both use the NaI Method
No Phenol No Chloroform

Wako

Wako BioProducts
Wako Chemicals USA, Inc.
1600 Bellwood Road
Richmond, VA 23237
Telephone: (804) 271-7677
Facsimile: (804) 271-7791
(800) 992-9256
bioproducts@wakousa.com

READER ENQUIRY NO. 107

Pepsyn Ltd
High Quality
Peptide Synthesis

'Standard' and 'non-standard'
Fast delivery
Competitive prices
Also advisory services

Contact: Dr J A Smith, Chairman, Pepsyn Ltd
School of Biological Sciences, Life Science Building
The University of Liverpool, Liverpool L69 7ZB, UK
Tel: 0151 794 4329 Fax: 0151 794 4349
e-mail: J.A.Smith@liverpool.ac.uk
Website: www.pepsyn.co.uk

READER ENQUIRY NO. 108

contact: info@TIB-MOLBIOL.de

TaqMan Probes...
TIB MOLBIOL

...and other fluorescent-labelled oligonucleotide probes for advanced nucleic acid detection. Dark quencher for all dual-labelled probes available.

TaqMan is a trademark of Hoffmann-La Roche. The use of the PCR and of the 5'-nuclease assay require a licence from Hoffmann-La Roche.

READER ENQUIRY NO. 109

Insulin assay

From ALPCO

www.alpco.com*A rat insulin ELISA for quantifying insulin in rat serum or plasma*

This assay, from American Laboratory Products Company (ALPCO), is a solid-phase two-site enzyme immunoassay available in a precoated strip format, 96 wells (12 × 8). It is based on the direct sandwich technique in which two monoclonal antibodies are directed against separate antigenic determinants on the insulin molecule. The measurement range is 0–5.5 $\mu\text{g l}^{-1}$, with an analytical detection limit of 0.07 $\mu\text{g l}^{-1}$. The kits come with ready-to-use liquid standards and the results are obtained within 2.5 hours.

Reader Service No. 110**Immunology reagents**

From Chemicon

www.chemicon.com*Antibodies, proteins and kits in the 1999 catalogue*

The Chemicon International Immunological Reagents catalogue contains over 1,000 new antibodies, proteins and kits for use in neuroscience, cell adhesion, extracellular matrix, cancer, cytokine, apoptosis, signal transduction, cytoskeletal, and infectious disease research. New products include antibodies to neurotransmitter transporters, synuclein, caspase assays, MMP activity assays, integrin antibody kits, cell cycle, ChemoKine cytokine ELISA kits, JAK/STAT pathway, surfactant proteins, aquaporins, western blot detection kits, antisense, primers and hybridization probes.

Reader Service No. 111**Phosphorylation specifics**

From Autogen Bioclear

www.autogen-bioclear.co.uk*Antibodies specific to a protein's phosphorylation state*

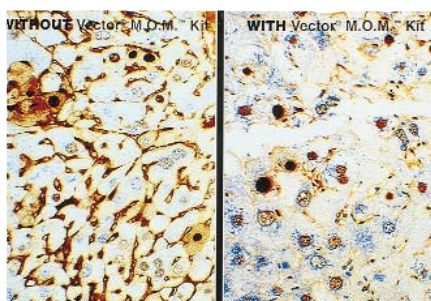
Autogen Bioclear produces a range of phosphorylation-state-specific antibodies. Included are antibodies to phosphorylated ERK (recognizes both ERK1 and ERK2), JNK (recognizes JNK1, JNK2 and JNK3), p38, p53, phosphotyrosine, GSK-3 β , JAK1 and JAK2 (both dual-phosphorylated), STAT1, STAT3 and C-Raf. Applications include western blotting, immunoprecipitation, immunohistochemistry and ELISA. Several are available as FITC, agarose and biotin conjugates.

Reader Service No. 112**Murine stain (1)**

From Vector

www.vectorlabs.com*Staining with mouse antibodies on mouse tissue sections*

The Vector Mouse On Mouse (M.O.M.) Immunodetection Kit offers a new method

**M.O.M. from Vector: before and after.**

for allowing specific staining on mouse tissue sections using mouse primary antibodies with, claims Vector, essentially no background staining in most tissues. A common problem associated with the use of mouse primary antibodies on mouse tissues is high background staining due to the inability of the secondary antibody to distinguish between the mouse primary antibody and endogenous mouse immunoglobulins in the tissue. The M.O.M. kit utilizes a novel blocking agent and a special detection methodology to eliminate background due to endogenous mouse immunoglobulins.

Reader Service No. 113**Protein blocking solution**

From Inotech

www.inotechintl.com*LB (Low Background) protein blocking solution for immunoassays*

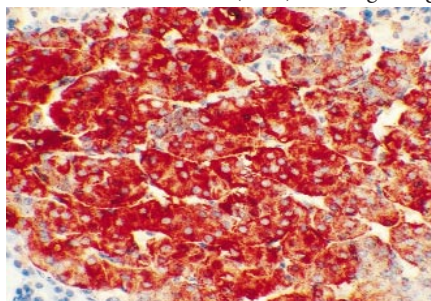
LB protein blocking solution is used to minimize nonspecific protein binding to assay surfaces, improving the specificity and sensitivity of immuno-binding assays and blotting procedures. It can also be used as a buffer additive (at 3 %) in immunoprecipitation reactions with protein A–Sepharose to enhance the yield. The solution is a liquid gelatin extract from cold water fish with properties similar to those of mammalian gelatin, but this gelatin stays liquid down to 4 °C, making it suitable for use in facilitating antigen–antibody reactions.

Reader Service No. 114**Murine stain (2)**

From InnoGenex

www.innogenex.com*Another mouse-to-mouse search*

The Mouse-on-Mouse Iso-IHC Kit offers immunohistochemical (IHC) staining using

**Mouse-on-mouse staining, this time InnoGenex**

mouse primary antibodies on mouse tissues. The detection system is claimed to give specific staining with no background. Any mouse antibody preparation can be used, since a proprietary biotin labelling reagent labels only antibodies, even in the presence of contaminating proteins. InnoGenex's Mouse Block reagent eliminates background staining due to endogenous tissue. The Iso-IHC Kits are available with a choice of AEC, BCIP/NBT, DAB, or Fast Red chromogen/substrate. Pictured (below, left) is mouse adrenal tissue stained with mouse anti-synaptophysin monoclonal.

Reader Service No. 115**Chromatography kits**

From Pierce

www.piercenet.com*Small-scale antibody purification made simpler*

Pierce Chemical Company's NAb Protein A, G and L Spin Chromatography Kits provide a convenient means for recovering or purifying a small amount of antibody (0.1 mg) from ascites or serum. The microcentrifuge tube and Spin X device, a bucket that fits inside the tube, provide an easy method for creating a small chromatography column and the smaller spin-tube format reduces protein loss. The kits work quickly, concentrating dilute amounts of antibody in one hour, and contain enough reagents for 10 purifications.

Reader Service No. 116**Protein G**

From Actigen

www.actigen.com*Protein G for affinity purification*

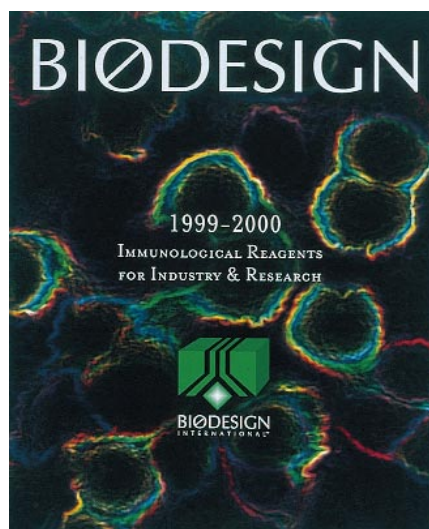
Actigen has now added Protein G to its family of affinity purification reagents that include rProtein L, rProtein LA and Protein A. Protein G is a bacterial protein from *Streptococcus* that binds to the Fc region of IgG antibodies. Protein G has similar binding characteristics to Protein A but binds more efficiently to bovine, goat, equine and sheep IgG. In addition, it adsorbs over a wider pH range. This well-tested affinity protein is provided coupled to agarose for purification, isolation and fractionation of IgG antibodies. It is available in 1- and 5-ml pack sizes and can also be provided in bulk at special rates. In addition, a horseradish peroxidase conjugate is available for detection of bound IgG antibody in assays and blotting.

Reader Service No. 117**Immunological catalogue**

From Biodesign

www.biodesign.com*Assay development reagents feature strongly*

The 1999–2000 catalogue, released on 1 May, features some 4,000 monoclonal and polyclonal antibodies, native and recombinant



Biodesign: apoptosis to viral antigens.

antigens and assay development reagents. Products headings include apolipoproteins, apoptosis, assay development, auto-immune, cytokines, immunoglobulins and viral antigens.

Reader Service No. 118

New antibodies

From RBI

www.callrbi.com

Tools for neuroscience research

Recent additions to the RBI product list include 28 antibodies for neuroscience research. This includes antibodies to protein kinase isoforms, specific subunits of the NMDA glutamate receptor, dopamine β -hydroxylase and synapsin-I. Antibodies are provided as monoclonals, polyclonal antisera and purified polyclonal antisera, and are available in liquid and lyophilized forms.

Reader Service No. 119

PAP Pen

From Research Products International

www.rpicorp.com

A marking pen designed for drawing thin film-like barriers on slides

The PAP Pen works by releasing quick-drying fluids in precise amounts to form a barrier with the proper surface tension needed to hold antibody solution and other specimens within a target area. It is effective



Anti-synapsin et al., from RBI.

for immunoperoxidase staining, ABC, frozen section and fluorescent antibody staining methods. The PAP Pen contains a water-repellent formulation that is insoluble in alcohol and acetone. The standard version is fitted with a 4-mm tapered point and contains enough fluid for over 800 applications. A Mini PAP Pen is available, with a 2.5 mm tip good for over 400 applications.

Reader Service No. 120

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

Check out our homepage at <http://www.resgen.com>

Peptides
\$14/residue
10 residue minimum

Polyclonal Antibodies - \$1,195

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

READER ENQUIRY NO. 14

Homogenizers?

- Tissue-Tearor™
- BeadBeater™
- Mini-BeadBeater™
- BioHomogenizer™

Check Us Out

www.biospec.com

BioSpec Products, Inc.
Ph/Fax 918-336-3363
e-mail info@biospec.com

New Probe Sizes 0.45cm!

READER ENQUIRY NO. 54

Diagnostic Reagents

Pefachrome®-Series: chromogenic and fluorogenic substrates

Pefabloc®-Series: serine proteinase inhibitors

Purified snake venom proteins (e.g. Protac®, Ecarin, Batroxobin)

PENTAPHARM LTD
CH 4002 Basel Fax: +41 61 319 96 19
<http://www.pentapharm.ch>

READER ENQUIRY NO. 16

Customised monoclonal & polyclonal antibodies with guarantee!

BioGenes GmbH
Köpenicker Str. 325 · D-12555 Berlin
Tel. +49-30-65 76 23 96 Fax: - 97
E-mail: BioGenes@Berlin.sireco.net
Internet: <http://www.biogenes.de>

READER ENQUIRY NO. 17

"Beside good science, there is much humour and much of interest outside the field of science. ... In my case, it kept me awake and delighted until the small hours"
Keith J. Laidler, Professor Emeritus of Chemistry, Ottawa University

The Bedside Book Genius and Eccentricity in Science 1869-1953

Edited by Walter Gratzer, with a foreword by Stephen Jay Gould

A fascinating selection of articles from Nature, offering a panorama of science seen against the background of 19th and 20th century history.



Price: £19.95/US\$29.95 + postage
Discount rate available for current subscribers.
To order, or for more information, contact: Nature Marketing
Tel. +44 (0) 171 843 4985
Fax. +44 (0) 171 843 4998
E-mail: subscriptions@nature.com